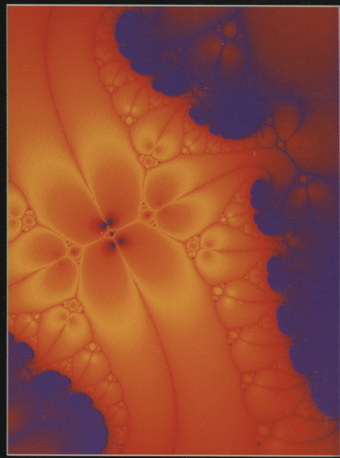
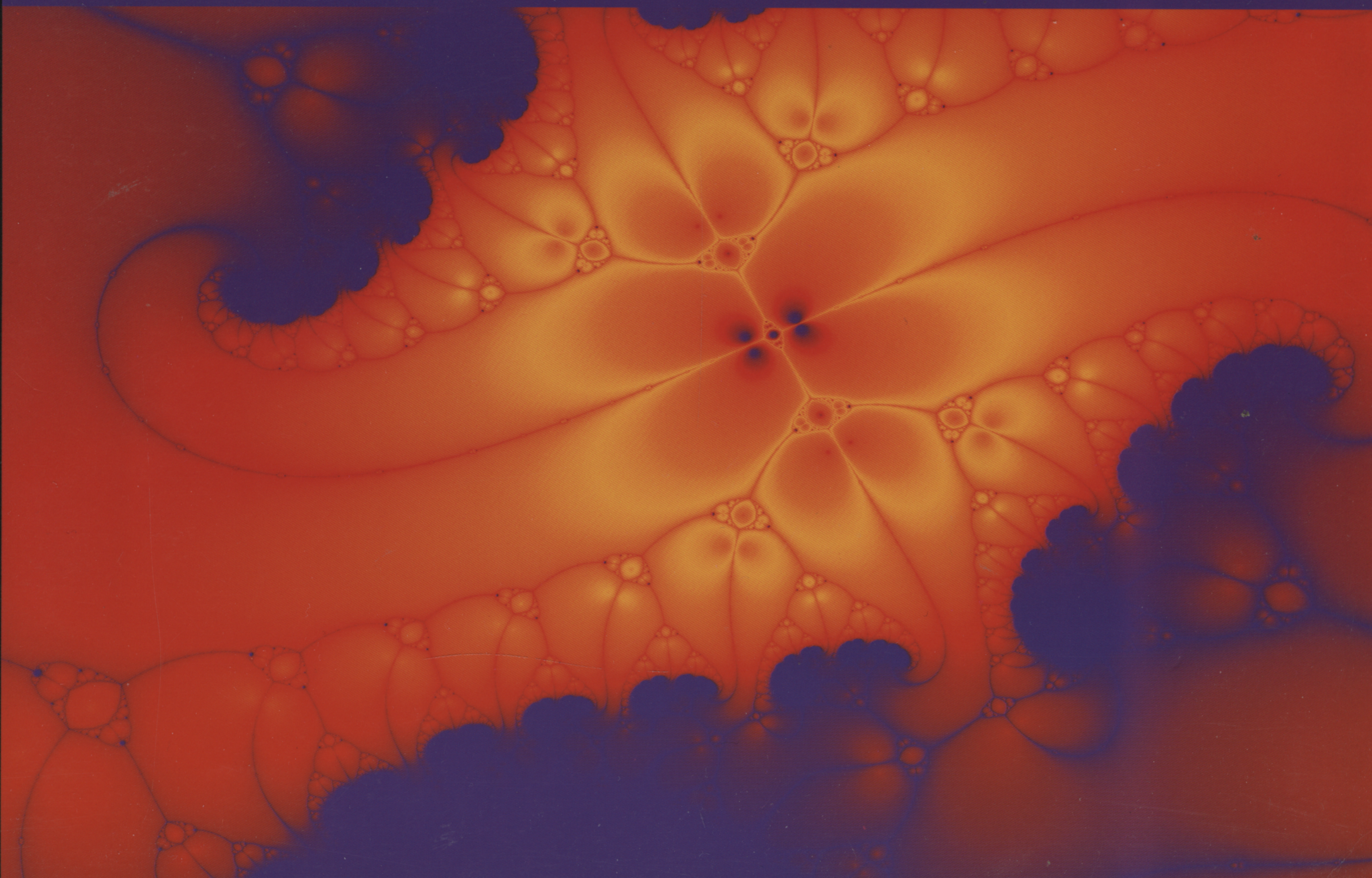
The Open University

M338 Topology



A1

Unit A1 Distance and continuity



The Open University

M338 Topology

A1

Distance and continuity

This publication forms part of an Open University course. Details of this and other Open University courses can be obtained from the Student Registration and Enquiry Service, The Open University, PO Box 197, Milton Keynes, MK7 6BJ, United Kingdom: tel. +44 (0)870 333 4340, e-mail general-enquiries@open.ac.uk

Alternatively, you may visit the Open University website at http://www.open.ac.uk where you can learn more about the wide range of courses and packs offered at all levels by The Open University.

To purchase a selection of Open University course materials, visit the webshop at www.ouw.co.uk, or contact Open University Worldwide, Michael Young Building, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom, for a brochure: tel. +44 (0)1908 858785, fax +44 (0)1908 858787, e-mail ouwenq@open.ac.uk

The Open University, Walton Hall, Milton Keynes, MK7 6AA.

First published 2006.

Edited, designed and typeset by The Open University, using the Open University TEX System.

Printed and bound in the United Kingdom by The Charlesworth Group, Wakefield.

ISBN 0 7492 4125 X

1.1

Contents

Introduction

This course provides an introduction to the mathematical topic of *topology*, which is the study of those properties of shapes that are unchanged by transformations that distort the shape but do not tear or stick together parts of the shape. Such properties of shapes are known as *topological invariants*. An example of a topological invariant is the number of holes an object possesses: if a transformation may not tear or stick, then the transformation cannot add holes to an object or remove holes from it.

At the start of our study of topology, there are various issues that we need to clarify. The first, and perhaps most important of these, is to specify what we mean by a transformation that does not tear or stick. The answer we obtain involves the notion of *continuity*.

This immediately leads to another problem: what is continuity? You are probably already familiar with the idea of a continuous function $f\colon \mathbb{R} \to \mathbb{R}$, from the set of real numbers to itself. For example, the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = x^3$$

is continuous as its graph has no jumps or breaks in it; whereas the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \frac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is *dis*continuous since there is a break in its graph at $x = 0$. However, in Block B we are interested in surfaces that lie in three-dimensional space, $\mathbb{R}^3$. What can it mean for a transformation between such surfaces to be continuous?

This unit addresses this problem by generalizing the definition of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ to functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. A crucial aspect of this process involves uncovering certain properties embedded in our idea of 'distance between points'. By using these properties we can extend the notion of 'distance' to more general spaces than $\mathbb{R}^n$, and this is the theme of later units in this block.

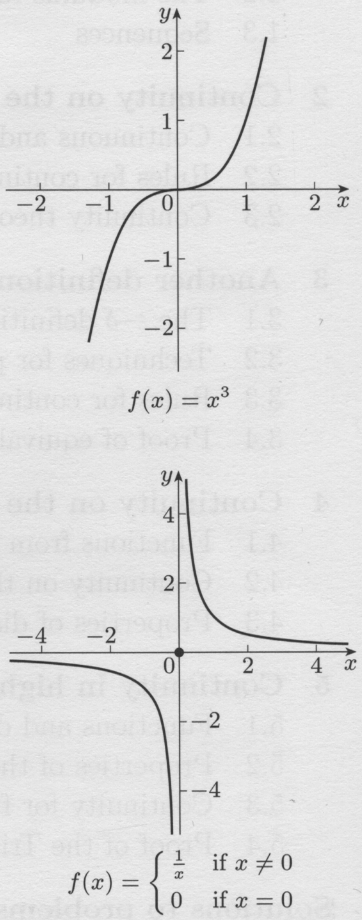


Figure 0.1

Many of the interesting properties possessed by continuous functions from $\mathbb{R}$ to $\mathbb{R}$ are also held by continuous functions in the more general setting, and these are the focus of Block C. The study of these properties has profound consequences for the development of modern mathematics.

Study guide

Before you study this unit, you should make sure that you have read the *Course Guide* and have looked through the first part of the *Handbook*, which gives a brief summary of the mathematical concepts with which you are expected to be familiar. You should refer back to this part of the *Handbook* when you need to remind yourself of a particular topic.

Also, before you attempt to use any of the computer software for the course, you will need to have worked through the *Computing Booklet*, in the process of which you will be expected to install the course software and familiarize yourself with the way it works. Since there is software associated with Section 3 of this unit, you might find it convenient to work through the *Computing Booklet* before you study this unit.

The DVD that accompanies this course provides an overview of topology. If you have not already watched it, you may like to do so now.

Although this is a long unit, many of the concepts and results should already be familiar to you. Also, much of the material in the last two sections consists of fairly straightforward generalizations of material earlier in the unit. You should, therefore, be able to work through the unit in rather less time than its length suggests.

Section 1, *Functions and sequences*, contains a reminder of the definition of a function and discusses one particular function — the modulus function — and various results that make use of this function. It also reminds you of the definition of a sequence and discusses various results concerning sequences. As much of this material may be familiar, you may wish to work through it quickly, simply to remind yourself of the basic concepts, results and notation.

In Section 2, *Continuity on the real line*, we begin our study of continuity by examining a definition for continuity for real functions phrased in terms of convergent sequences. Again, this material may be already familiar to you, and so you may wish to skim through it simply to confirm that you know the basic concepts, methods and techniques.

Section 3, *Another definition of continuity*, concerns an alternative definition of continuity which is the starting point for future development. Even if this material is familiar, you should study this section carefully since it contains several important ideas. To assist you in understanding this section, there is some associated software. Subsection 3.4 is not assessed and, if you are short of time, you may prefer to omit it on a first reading.

In Section 4, *Continuity on the plane*, we extend the definition of continuity to real-valued functions on the plane, $\mathbb{R}^2$. This is where we start our investigations into the idea of distance; from them we shall be able to develop a more general definition of continuity.

In Section 5, *Continuity in higher dimensions*, we continue this process of extension by discussing the continuity of functions between any two Euclidean spaces. This section contains a number of important ideas that provide a foundation for future units. Subsection 5.4 is not assessed and, if you are short of time, you may prefer to omit it on a first reading.

1 Functions and sequences

After working through this section, you should be able to:

- state the definition of a *function*;
- use the *modulus function* in inequalities and obtain the solution sets of inequalities involving the modulus function;
- use the *Triangle Inequality* and its variants in $\mathbb{R}$;
- explain the meanings of *sequence, null sequence, convergent sequence, divergent sequence* and *subsequence*;
- test a given sequence for convergence or divergence.

We start our investigation of continuity by clarifying some of the important terminology and notions on which the definition of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ is based.

We assume that you are familiar with the idea of a set and some basic set theory. We also assume that you are familiar with the following sets of numbers:

A brief summary of the concepts with which you should be familiar is given in the first part of the *Handbook*.

- the **natural numbers** $\mathbb{N} = \{1, 2, 3, \ldots\}$;
- the **integers** $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$;
- the **rational numbers** $\mathbb{Q} = \{p/q : p \in \mathbb{Z},\ q \in \mathbb{N}\}$;
- the **real numbers** $\mathbb{R}$, which can be represented by infinite decimals;
- the **irrational numbers** $\mathbb{R} - \mathbb{Q}$, which are those real numbers that are not rational, such as $\sqrt{2}$, π and e.

For two sets A and B, $A - B$ denotes the set of elements of A that are not in B. Thus $\mathbb{R} - \mathbb{Q}$ denotes the set of real numbers that are not rational — that is, the irrationals. There is no standard single-letter symbol for the irrationals.

We assume that you are familiar with the usual arithmetic operations of addition, subtraction, multiplication and division of real numbers, with the different sorts of interval on the real line, and with manipulating simple inequalities involving real numbers. We also assume a basic familiarity with the trigonometric and exponential functions.

1.1 Functions between sets

Although in Sections 1–3 of this unit we are concerned with functions from $\mathbb{R}$ to $\mathbb{R}$, we first give the general definition of a *function* between two *arbitrary* sets. This has the advantage that we can use the same definition each time we change to a new setting.

Functions are also called *mappings*, or *maps*. In some texts, the words 'mapping' and 'map' denote only continuous functions, but this is not our usage.

Definition

A **function** f is defined by specifying:

- a set A, the **domain** of f;
- a set B, the **codomain** of f;
- a **rule** $x \mapsto f(x)$ that associates with each element $x \in A$ a *unique* element $f(x) \in B$.

The element $f(x)$ is the **image** of x under f and the set $f(A) = \{f(x) : x \in A\} \subseteq B$ is the **image set** of A under f.

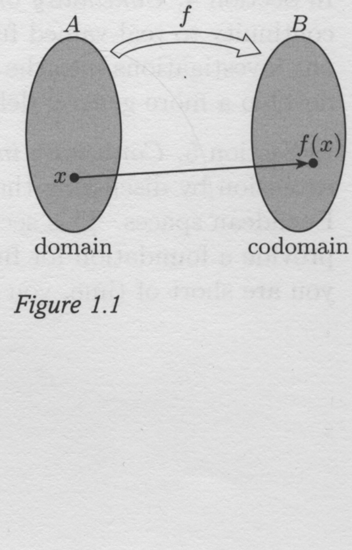


Figure 1.1

Remark

When writing functions symbolically we use one of three forms. For example, for the function from $\mathbb{R}$ to $\mathbb{R}$ given by $f(x) = x^2$, we use

$$\begin{array}{lll} \text{either} & f\colon \mathbb{R} \to \mathbb{R}, & f\colon x \mapsto x^2, \\ \text{or} & f\colon \mathbb{R} \to \mathbb{R}, & f(x) = x^2, \\ \text{or} & f(x) = x^2 & (x \in \mathbb{R}). \end{array}$$

In the first case we may sometimes simply write $x \mapsto x^2$ rather than $f\colon x \mapsto x^2$.

Problem 1.1

Which of the following rules define functions with domain $\mathbb{Z}$ and codomain $\mathbb{N}$?

(a) Square each integer and add 1.

(b) Associate with each integer the set of all its divisors.

(c) Associate with each integer the number of its divisors.

Recall that the *divisors* of an integer z are the natural numbers n such that z/n is an integer.

1.2 The modulus function

This subsection discusses the *modulus function* — a function that will prove vital in our study of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$. We shall also derive some important results that make use of the modulus function.

The **absolute value** $|x|$ of a real number x is defined by

$|x|$ is often called the *modulus of* x.

$$|x| = \begin{cases} +x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$$

For example, $|3| = 3$ and $|-3| = 3$.

Definition

The **modulus function** is the function

$$f\colon \mathbb{R} \to \mathbb{R}, \quad f\colon x \mapsto |x|.$$

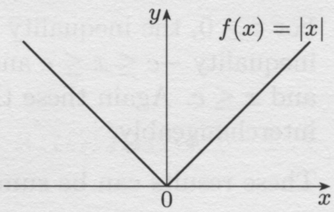


Figure 1.2

The graph of the modulus function is shown in Figure 1.2.

For each real number a,

$$0 \leq |a|, \quad a \leq |a|, \quad -|a| \leq a, \quad |-a| = |a|. \tag{1.1}$$

For example, when $a = -2$,

$$0 \leq 2 = |-2|, \quad -2 \leq 2 = |-2|, \quad -|-2| = -2 \leq -2, \quad |-(-2)| = 2 = |-2|.$$

Geometrically, $|a|$ gives the *distance*, without regard to direction, along the real line from a to 0.

For any real numbers a and b,

$$|ab| = |a|\,|b|, \tag{1.2}$$

$$|a|^2 = a^2. \tag{1.3}$$

For example, when $a = -2$ and $b = 3$,

$$|-2 \times 3| = |-6| = 6 = 2 \times 3 = |-2| \times |3|,$$
$$|-2|^2 = 2^2 = 4 = (-2)^2.$$

Note also that $|a| = 0$ if and only if $a = 0$.

One of the main uses of the modulus function is in *inequalities*, so it is useful to look at the solutions x to the inequality

$$|x| < c,$$

where c is a constant in $\mathbb{R}$. The solution set depends on whether $c > 0$ or $c \leq 0$. For example, for $c = 2$, the set of x for which $|x| < c$ is the *open interval* $(-2, 2) = \{x : -2 < x < 2\} = \{x : |x| < 2\}$, as Figure 1.3 illustrates. However, for $c = -2$, there is no number x for which $|x| < c$, and so the solution set is the empty set $\varnothing$. In general, for $c > 0$ the solution set is the open interval $(-c, c) = \{x : -c < x < c\}$ and for $c \leq 0$ the solution set is the empty set $\varnothing$.

In some texts the open interval is denoted by $]-2, 2[$.

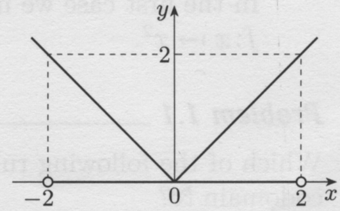


Figure 1.3

For $c > 0$, the condition that

$$|x| < c$$

is equivalent to the double-sided inequality

$$-c < x < c.$$

This says, informally, that x lies between $-c$ and $+c$ on the real line: in other words, x is within a distance c of 0, as Figure 1.4 illustrates. The condition $-c < x < c$ is equivalent to saying that x satisfies the two simultaneous inequalities

$$-c < x \quad \text{and} \quad x < c.$$

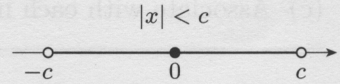


Figure 1.4

For $c > 0$, the equivalent notations $|x| < c$, $-c < x < c$, and $-c < x$ and $x < c$ will be used interchangeably.

We have similar results when the inequality is

$$|x| \leq c.$$

For $c \geq 0$, the solution set is the *closed interval* $[-c, c] = \{x : -c \leq x \leq c\}$. For $c < 0$, the solution set is the empty set $\varnothing$.

When $c = 0$, the solution set is $[0, 0]$, which is simply the one-point set $\{0\}$.

For $c \geq 0$, the inequality $|x| \leq c$ is equivalent both to the double-sided inequality $-c \leq x \leq c$ and to the pair of simultaneous inequalities $-c \leq x$ and $x \leq c$. Again these three equivalent notations will be used interchangeably.

These results can be summarized as follows.

Solution sets of inequalities 1

- For $c > 0$, the inequalities
  (a) $|x| < c$, (b) $-c < x < c$, (c) $-c < x$ and $x < c$
  are equivalent. Moreover, for $c \in \mathbb{R}$,
  $$\{x \in \mathbb{R} : |x| < c\} = \begin{cases} (-c, c) & \text{if } c > 0, \\ \varnothing & \text{if } c \leq 0. \end{cases}$$
- For $c \geq 0$, the inequalities
  (a) $|x| \leq c$, (b) $-c \leq x \leq c$, (c) $-c \leq x$ and $x \leq c$
  are equivalent. Moreover, for $c \in \mathbb{R}$,
  $$\{x \in \mathbb{R} : |x| \leq c\} = \begin{cases} [-c, c] & \text{if } c \geq 0, \\ \varnothing & \text{if } c < 0. \end{cases}$$

Problem 1.2

Find the solution set of each of the following inequalities:

(a) $|x| < 3$; (b) $|x| \leq -3$; (c) $|3x| < 3$.

We now consider more complicated inequalities.

Worked problem 1.1

For real numbers b and c, with $c > 0$, find the set of all real numbers x for which $|x - b| < c$.

Solution

The condition $|x - b| < c$ is equivalent to

$$-c < x - b < c,$$

which is the case if and only if

$$b - c < x < b + c.$$

Thus,

$$\{x : |x - b| < c\} = (b - c, b + c).$$

So the required set is the open interval around b consisting of those points within a distance c of b. ■

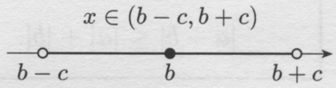


Figure 1.5

A similar result holds for the inequality $|x - b| \leq c$, and so we have the following general results.

> ***Solution sets of inequalities 2***
>
> ▶ For $b, c \in \mathbb{R}$ with $c > 0$,
>
> $$\{x \in \mathbb{R} : |x - b| < c\} = (b - c, b + c).$$
>
> ▶ For $b, c \in \mathbb{R}$ with $c \geq 0$,
>
> $$\{x \in \mathbb{R} : |x - b| \leq c\} = [b - c, b + c].$$

Problem 1.3

Find the set of all numbers x for which:

(a) $|x - 1| < 3$; (b) $|x - 1| < 3$ and $|x + 1| < 3$.

In general, for any real numbers a and b,

$$|b - a| = |a - b|, \tag{1.4}$$

and this quantity gives the *distance* between a and b along the real line, as Figure 1.6 illustrates. Notice that $|b - a| = 0$ if and only if $a = b$.

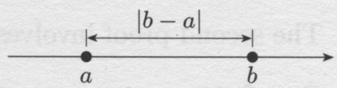


Figure 1.6

The following problems concern general properties of inequalities involving the modulus function.

Worked problem 1.2

Prove that, for real numbers a and b,

$$|a| \leq |b| \quad \text{if and only if} \quad -|b| \leq a \leq |b|.$$

Solution

For $c \geq 0$, we have the equivalent expressions

$$|x| \leq c \quad \text{if and only if} \quad -c \leq x \leq c.$$

Setting $x = a$ and $c = |b|$, we obtain

$$|a| \leq |b| \quad \text{if and only if} \quad -|b| \leq a \leq |b|. \qquad \blacksquare$$

Problem 1.4

Prove that, for real numbers a and b,

$$|a| \leq |b| \quad \text{if and only if} \quad -|b| \leq |a| \leq |b|.$$

Another important property of the modulus function is the *Triangle Inequality*.

> ***Theorem 1.1 Triangle Inequality***
>
> For all real numbers a and b,
>
> $$|a+b| \leq |a| + |b|.$$

For example, with $a = -4$ and $b = 3$, $|-4+3| = |-1| = 1$ and $|-4| + |3| = 4 + 3 = 7$, so $|-4+3| \leq |-4| + |3|$.

Problem 1.5

Verify the Triangle Inequality when:
(a) $a = 3, b = 4$; (b) $a = -3, b = 4$; (c) $a = -3, b = -4$.

There are several ways to prove the Triangle Inequality. We give two proofs. The first involves squaring both sides.

Proof For all real numbers a and b, we have

$$\begin{aligned} |a+b|^2 &= (a+b)^2 \qquad \text{(by (1.3))} \\ &= a^2 + 2ab + b^2 \\ &\leq a^2 + 2|ab| + b^2 \qquad \text{(by (1.1))} \\ &= |a|^2 + 2\,|a|\,|b| + |b|^2 \qquad \text{(by (1.2) and (1.3))} \\ &= (|a| + |b|)^2. \end{aligned}$$

We now use the fact that, when p and q are non-negative numbers, $p^2 \leq q^2$ implies $p \leq q$. Hence

$$|a+b| \leq |a| + |b|. \qquad \blacksquare$$

The second proof involves the manipulation of inequalities.

Proof If a and b are real numbers, then from (1.1)

$$-|a| \leq a \leq |a| \quad \text{and} \quad -|b| \leq b \leq |b|.$$

Adding these two double-sided inequalities together, we obtain

$$-|a| - |b| \leq a + b \leq |a| + |b|,$$

which can be written as

$$-(|a| + |b|) \leq a + b \leq |a| + |b|.$$

This is equivalent to

$$|a+b| \leq |a| + |b|. \qquad \blacksquare$$

Problem 1.6

Use the Triangle Inequality to show that:

(a) for all real numbers a, b and c, $|a+b+c| \leq |a| + |b| + |c|$;

(b) for all real numbers a and b, $|a-b| \leq |a| + |b|$.

Problem 1.7

Show that the Triangle Inequality is equivalent to the inequality

$$|c-a| \leq |b-a| + |c-b|,$$

where a, b and c are any real numbers.

Hint To show equivalence, you will need to show first that the Triangle Inequality implies the given inequality and then the converse. To show the former, write the Triangle Inequality as $|p+q| \leq |p| + |q|$, for all real numbers p and q, and then write p and q in terms of a, b and c. A similar technique will allow you to prove the converse.

The following theorem provides a useful variant of the Triangle Inequality, known as the *reversed form of the Triangle Inequality* — or, more briefly, as the *Reverse Triangle Inequality.*

Theorem 1.2 Reverse Triangle Inequality

For all real numbers a and b,

$$|b-a| \geq \big||b| - |a|\big|.$$

We prove this theorem in a more general context in Section 5.

1.3 Sequences

Ever since learning to count, you have been familiar with the sequence of natural numbers:

$$1, 2, 3, 4, 5, \ldots .$$

You will also have met other sequences of numbers — for example,

$$2, 4, 6, 8, 10, \ldots \quad \text{and} \quad \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{16}, \tfrac{1}{32}, \ldots .$$

Definition

A **(real) sequence** is an unending ordered list of real numbers

$$a_1, a_2, a_3, \ldots .$$

The real number a_n ($n \in \mathbb{N}$) is the ***n*th term** of the sequence, and the whole sequence is denoted by (a_n).

Some texts use the notation $\{a_n\}$ instead of (a_n).

Remarks

(i) Alternatives to the notation (a_n) which we shall sometimes use are $(a_n)_{n=1}^{\infty}$, $(a_n)_{n\in\mathbb{N}}$ and $\mathbf{a}$.

(ii) The terms of a sequence (a_n) do not need to be different. For example, $-1, 1, -1, 1, \ldots$ and $1, 1, 1, 1, \ldots$ are sequences. A sequence all of whose terms are the same is known as a **constant sequence**.

(iii) For any subset A of $\mathbb{R}$, we say that a sequence (a_n) is **in** A if $a_n \in A$ for all n. For example, the sequences $2, 4, 6, 8, \ldots$ and $1, 1, 1, 1, \ldots$ are in $\mathbb{N}$.

(iv) A sequence can also be viewed as a function. A formal definition is:

a *(real) sequence* is a function $a: \mathbb{N} \to \mathbb{R}$ given by $n \mapsto a_n$.

We often define sequences by giving the first few terms and a formula for the nth term — for example,

$$2, 4, 6, \ldots, 2n, \ldots .$$

This sequence may also be written as

$$a_n = 2n \text{ for } n = 1, 2, 3, \ldots$$

or as $(2n)_{n=1}^{\infty}$ or simply as $(2n)$.

Sometimes the nth term is defined by a recursion formula, as in the sequence

$$1, 1, 2, 3, 5, 8, \ldots, \text{ where } a_n = a_{n-1} + a_{n-2} \text{ for } n \geq 3.$$

If you look at the terms of the sequence

$$1, \tfrac{1}{2}, \tfrac{1}{3}, \ldots, \tfrac{1}{n}, \ldots$$

you will notice that, as n gets large, the terms get closer to 0. This sequence is an example from an important class of sequences, the class of *null sequences* — the sequences that *converge to* 0.

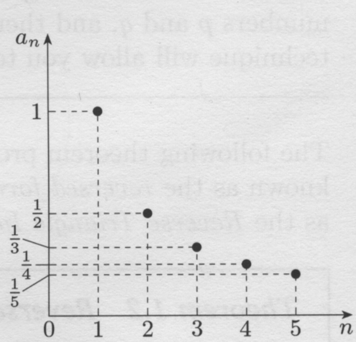


Figure 1.7 The first few terms of the sequence $(1/n)$

This notion of converging to 0 can be made precise. The idea is to show that, for *any* positive number ε, we can find a natural number N such that all of the terms of the sequence after a_N lie within the interval $(-\varepsilon, \varepsilon)$. A sequence is a null sequence if this property holds whatever positive number ε is taken, no matter how small. For example, with the sequence $(1/n)$, suppose that $\varepsilon = 0.01$. If $(1/n)$ is a null sequence then there must be a value N for n for which all of the subsequent terms in the sequence are within 0.01 of 0. In fact, this occurs when $N = 100$ because, from $n = 101$ onwards, $1/n$ is smaller than 0.01.

Definition

A sequence (a_n) is a **null sequence** if, for all $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$|a_n| < \varepsilon \quad \text{whenever } n > N.$$

We say that a null sequence **converges to 0**, and write

$$a_n \to 0 \text{ as } n \to \infty, \qquad \text{or simply } a_n \to 0.$$

Often, we say simply that '(a_n) is null'.

Remarks

(i) In general, for a given sequence, the value of N depends on ε; usually, the smaller ε is, the larger N will be.

(ii) The definition uses the absolute value $|a_n|$ because terms of the sequence may be negative. As you will recall from Subsection 1.2, $|a_n| < \varepsilon$ means that $-\varepsilon < a_n < \varepsilon$: that is, on the graph of a_n against n, the terms of a null sequence eventually lie inside a band of width 2ε about 0, as Figure 1.8 illustrates.

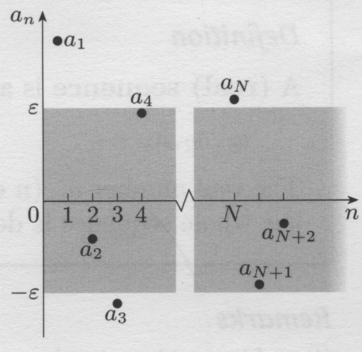


Figure 1.8

Worked problem 1.3

Prove that each of the following sequences is null:

(a) (a_n), where $a_n = 0$ for all n;

(b) (b_n), where $b_n = 1/n^2$ for all n;

(c) (c_n), where $c_n = (-1)^n/n^2$ for all n.

Solution

(a) Let $\varepsilon > 0$ be given. Then $|a_n| = |0| = 0 < \varepsilon$ for all n.

Choose $N = 1$. Then $|a_n| < \varepsilon$ whenever $n > N$.

We deduce that (a_n) is null.

(b) Let $\varepsilon > 0$ be given. In this case $|b_n| = |1/n^2| = 1/n^2$ and

$$1/n^2 < \varepsilon \quad \text{if and only if} \quad 1/\sqrt{\varepsilon} < n.$$

Choose $N = [1/\sqrt{\varepsilon}]$, the largest integer less than or equal to $1/\sqrt{\varepsilon}$. Then, for all $n > N$, we have $n > 1/\sqrt{\varepsilon}$, and hence

Here, $[x]$ denotes the largest integer less than or equal to x: for example, $[2] = 2$, $[1.23] = 1$ and $[-1.5] = -2$.

$$|b_n| = 1/n^2 < \varepsilon \quad \text{whenever } n > N.$$

We deduce that (b_n) is null.

(c) Let $\varepsilon > 0$ be given. Then, $|c_n| = |(-1)^n/n^2| = 1/n^2 = b_n$ for all n.

Hence we choose N as in (b), so that $|c_n| = |b_n| < \varepsilon$ whenever $n > N$.

We deduce that (c_n) is null. ■

Problem 1.8

Prove that the sequence (a_n) given by $a_n = 1/\sqrt{n}$ is null.

To prove that a sequence is null, we have to show that, whatever band of width 2ε we make around 0, the terms of the sequence must eventually all lie inside that band. Thus to prove that a sequence (a_n) is *not* a null sequence, we need to show that this condition does not hold. This means that we need to find *just one* band around 0 for which, no matter how far we progress in the sequence, there are always later terms of the sequence that lie outside that band. More formally, we need to find one value of ε such that, no matter what choice of N we make, we can always find a natural number $n > N$ for which $|a_n| \geq \varepsilon$. To put it slightly differently, we need to find an ε for which there is *no* natural number N such that $|a_n| < \varepsilon$ whenever $n > N$.

Worked problem 1.4

Prove that the sequence (a_n) given by $a_n = (-1)^n$ is not null.

Solution

Since the sequence is $-1, 1, -1, 1, \ldots$ we can take any $\varepsilon < 1$, say $\varepsilon = \frac{1}{2}$. Then, since $|a_n| = 1$ for all n, $|a_n| \geq \varepsilon$ for all n. So there is no natural number N for which $|a_n| < \varepsilon$ for all $n > N$. Thus the sequence is not null. ■

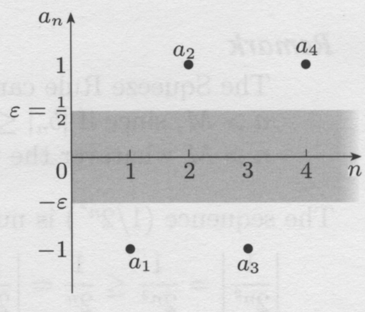


Figure 1.9

Problem 1.9

Prove that the constant sequence (a_n) given by $a_n = 2$ for all n is not null.

Here are some basic null sequences.

We shall not prove that these sequences are null.

Basic null sequences

The following sequences are null:

- $(1/n^p)$, for any constant $p > 0$;
- $(n^p c^n)$, for any constants $p \in \mathbb{R}$ and $|c| < 1$;
- $(c^n/n!)$, for any constant $c \in \mathbb{R}$;
- $(n^p/n!)$, for any constant $p \in \mathbb{R}$.

Examples:
$(1/n^3)$ when $p = 3$;
$(1/2^n)$ when $c = \frac{1}{2}$, $p = 0$;
$(2^n/n!)$ when $c = 2$;
$(n^{100}/n!)$ when $p = 100$.

It is often tricky to prove directly from the definition that a given sequence is null. However, there are several rules that allow us to combine null sequences to form new null sequences. They are summarized as follows.

Combination Rules for null sequences

Sum Rule If (a_n) and (b_n) are null, then $(a_n + b_n)$ is null.

Multiple Rule If (a_n) is null, then (λa_n) is null for any $\lambda \in \mathbb{R}$.

Product Rule If (a_n) and (b_n) are null, then $(a_n b_n)$ is null.

For example, since $(1/n)$ and $(1/\sqrt{n})$ are (basic) null sequences, the Sum Rule implies that $(1/n + 1/\sqrt{n})$ is a null sequence, and the Product Rule implies that $(1/n^{3/2})$ is a null sequence.

Alternatively, we could note from our list of basic null sequences that $(1/n^{3/2})$ is a sequence of the type $(1/n^p)$.

Problem 1.10

Prove the Multiple Rule for null sequences: if (a_n) is null, then (λa_n) is null for any $\lambda \in \mathbb{R}$.

We ask you to prove the other Combination Rules for null sequences in the problems for this unit.

Another useful rule for proving that sequences are null is the *Squeeze Rule.* This rule uses the idea that, given a known null sequence, a sequence whose terms are eventually closer to 0 than the corresponding terms of the given sequence must also be null.

Squeeze Rule If (a_n) is null and if there is an $M \in \mathbb{N}$ such that $|b_n| \leq |a_n|$ for all $n > M$, then (b_n) is null.

Remark

The Squeeze Rule can often be applied for all $n \in \mathbb{N}$ rather than for $n > M$, since if $|b_n| \leq |a_n|$ holds for all $n \in \mathbb{N}$ then it must hold for all $n > M$ whatever the value of $M \in \mathbb{N}$.

The sequence $(1/2^{n^2})$ is null. This follows from the Squeeze Rule, since

$$\left|\frac{1}{2^{n^2}}\right| = \frac{1}{2^{n^2}} \leq \frac{1}{2^n} = \left|\frac{1}{2^n}\right| \quad \text{for all } n \in \mathbb{N}$$

and $(1/2^n)$ is a basic null sequence.

A proof of the Squeeze Rule follows.

Proof Let (a_n) be null, and suppose that there is an $M \in \mathbb{N}$ such that $|b_n| \leq |a_n|$ for all $n > M$. Let $\varepsilon > 0$ be given. Since (a_n) is null, there is an $N \in \mathbb{N}$ such that $|a_n| < \varepsilon$ whenever $n > N$. Let $K = \max\{M, N\}$. It follows that

Recall that $\max\{M, N\}$ denotes the greater of M and N.

$$|b_n| \leq |a_n| < \varepsilon \quad \text{whenever } n > K,$$

and so (b_n) is also null. ■

Worked problem 1.5

Let (a_n) be a null sequence and let $(b_n) = (a_n^2/(1 + |a_n|))$.
Use the Combination and Squeeze Rules to prove that (b_n) is null.

Solution

We first observe that, by the Product Rule, (a_n^2) is null since it is the product of a_n with itself.

Now notice that, since $|a_n| \geq 0$ for all n,

$$|1/(1 + |a_n|)| \leq 1$$

for all n, and so

$$|a_n^2/(1 + |a_n|)| \leq |a_n^2|$$

for all n. Hence, by the Squeeze Rule, $(b_n) = (a_n^2/(1 + |a_n|))$ is null. ■

Problem 1.11

Let (a_n) be a null sequence, let $\lambda \in \mathbb{R}$ with $\lambda \geq 0$ and let (b_n) be a sequence such that $|b_n| \leq \lambda |a_n|$ for all n. Use the Combination and Squeeze Rules to prove that (b_n) is null.

We can use null sequences to define the convergence of sequences to values other than 0.

Definition

A sequence (a_n) **converges** to $l \in \mathbb{R}$ if the sequence $(a_n - l)$ is a null sequence: that is, if, for all $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$|a_n - l| < \varepsilon \quad \text{whenever } n > N.$$

We write $a_n \to l$ as $n \to \infty$, or simply $a_n \to l$.

A sequence that does not converge to any real number is **divergent**.

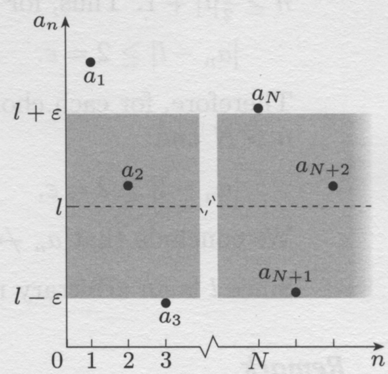


Figure 1.10

Remarks

(i) A sequence converges to l if, for each $\varepsilon > 0$, the terms of the sequence eventually lie inside a band of width 2ε about l, as Figure 1.10 illustrates.

(ii) If a sequence (a_n) does not converge to l, then we write $a_n \not\to l$ as $n \to \infty$. This means that either the sequence is divergent or it converges to some number other than l.

Problem 1.12

State conditions involving ε and N for a sequence (a_n) to be divergent, i.e. state conditions such that (a_n) does not converge to *any* limit l.

Hint Use the discussion, following Problem 1.8, of the conditions for showing that a sequence is not a null sequence.

Worked problem 1.6

Prove that:

(a) the constant sequence $5, 5, 5, \ldots$ converges to 5;

(b) the sequence $2, 4, 6, \ldots, 2n, \ldots$ is divergent.

Solution

(a) Let (a_n) be the constant sequence $5, 5, 5, \ldots$, and consider the sequence (b_n), where $b_n = a_n - 5$ for all n.

Then $b_n = 0$ for all n, and so is the constant sequence considered in Worked problem 1.3(a). Hence, $(b_n) = (a_n - 5)$ is a null sequence, and so $a_n \to 5$ as $n \to \infty$.

(b) Here, $a_n = 2n$ for all n.

To prove divergence, we have to show that, given any l, there is an $\varepsilon > 0$ such that, no matter what choice of N we make, we can always find $n > N$ for which $|a_n - l| \geq \varepsilon$.

To show divergence, it is sufficient to find just *one* such $\varepsilon > 0$.

We take an arbitrary value of l and use the Reverse Triangle Inequality (Theorem 1.2) to find a lower bound on the size of $|a_n - l|$:

$$|a_n - l| = |2n - l| \geq 2n - |l|. \tag{1.5}$$

Let us take $\varepsilon = 2$. Thus, we shall have demonstrated divergence if we can show that, for any $N \in \mathbb{N}$, there is an $n > N$ for which $|a_n - l| \geq 2$, and (1.5) tells us that this will be true whenever

In this case, we could choose $\varepsilon = 1$ or 0.5 or 7, or any positive value! We choose $\varepsilon = 2$ to make the algebra easier.

$$2n - |l| \geq 2.$$

This inequality holds when $2n \geq |l| + 2$, or equivalently when $n \geq \frac{1}{2}|l| + 1$. Thus, for all $n \geq \frac{1}{2}|l| + 1$, we have

$$|a_n - l| \geq 2 = \varepsilon.$$

Therefore, for each choice of N, if we take $n > \max\{N, \frac{1}{2}|l| + 1\}$ then $n > N$ and

$$|a_n - l| \geq 2 = \varepsilon.$$

We conclude that $a_n \not\to l$ as $n \to \infty$.

Since l is an arbitrary real number, (a_n) is a divergent sequence. ■

Remark

Any constant sequence $l, l, l, \ldots$ converges to its constant value l.

As with null sequences, there are Combination Rules for combining convergent sequences to form new convergent sequences.

> ***Combination Rules for convergent sequences***
>
> If $a_n \to l$ and $b_n \to m$ as $n \to \infty$, then:
>
> **Sum Rule** $a_n + b_n \to l + m$ as $n \to \infty$;
>
> **Multiple Rule** $\lambda a_n \to \lambda l$ as $n \to \infty$, for any $\lambda \in \mathbb{R}$;
>
> **Product Rule** $a_n b_n \to lm$ as $n \to \infty$;
>
> **Quotient Rule** $a_n / b_n \to l/m$ as $n \to \infty$, provided that $m \neq 0$ and $b_n \neq 0$ for all n.

These rules can be used, together with the basic null sequences, to prove that other sequences are convergent.

Worked problem 1.7

Prove that the following sequences converge, and find their limits.

(a) $(a_n) = \left((6 + 2^{-n})\left(2 - \frac{1}{n}\right)\right)$ (b) $(b_n) = \left(\dfrac{3n! + 2^n}{n^{50} + n!}\right)$

Solution

(a) Both (2^{-n}) and $(\frac{1}{n})$ are basic null sequences. Furthermore, by the Multiple Rule for null sequences (with $\lambda = -1$), the sequence $(-\frac{1}{n})$ is null. Also (6) and (2) are constant sequences and so converge to 6 and 2 respectively.

Hence, by the Sum Rule,

$(6 + 2^{-n})$ converges to $6 + 0 = 6$ and $(2 - \frac{1}{n})$ converges to $2 + 0 = 2$.

Then, by the Product Rule,

$(a_n) = ((6 + 2^{-n})(2 - \frac{1}{n}))$ converges to $6 \times 2 = 12$.

(b) Since $n!$ is the dominant term of b_n (that is, it grows faster than 2^n or n^{50} as n increases), we divide both the numerator and the denominator by $n!$. This gives

$$b_n = \frac{3 + 2^n/n!}{n^{50}/n! + 1}.$$

Both $(2^n/n!)$ and $(n^{50}/n!)$ are basic null sequences, and (3) and (1) are constant sequences. So, by the Sum Rule, the sequence $(3 + 2^n/n!)$ converges to 3 and the sequence $(n^{50}/n! + 1)$ converges to 1. Hence, by the Quotient Rule, (b_n) converges to $3/1 = 3$. ■

Remark

For sequences whose general term is expressed as a quotient, the technique of dividing both the numerator and the denominator by the *dominant term* can often be useful when trying to prove convergence or divergence.

For constants c and p, $n!$ dominates c^n, which in turn dominates n^p.

Problem 1.13

Prove that the following sequence converges, and find its limit.

$$\left(\frac{4n^2 + 1000^n}{n! + 7n^3}\right)$$

A useful idea for examining the convergence of sequences is that of *subsequences*. For example, two subsequences of the sequence

$$1, 2, 3, 4, 5, \ldots, n, \ldots$$

are

$$1, 3, 5, 7, 9, \ldots, 2n-1, \ldots \quad \text{and} \quad 1, 8, 27, 64, 125, \ldots, n^3, \ldots.$$

The essential features of a subsequence are that all its terms are in the original sequence, and that the terms occur in the same order as in the original sequence. In general, given a sequence (a_n), we can form many different subsequences.

Definition

The sequence $(a_{n_k})_{k=1}^{\infty}$ is a **subsequence** of the sequence (a_n) if $(n_k)_{k=1}^{\infty}$ is a strictly increasing sequence of positive integers, i.e. if $1 \leq n_1 < n_2 < n_3 < \ldots$.

In particular:
when $n_k = 2k$, we have $(a_{2k})_{k=1}^{\infty}$, which is the **even subsequence** of (a_n);
when $n_k = 2k-1$, we have $(a_{2k-1})_{k=1}^{\infty}$, which is the **odd subsequence** of (a_n).

Remarks

(i) Every sequence is a subsequence of itself, obtained by putting $n_k = k$, for $k = 1, 2, 3, \ldots$.

(ii) The even subsequence (a_{2k}) consists of the even-numbered terms $a_2, a_4, a_6, \ldots$ of the original sequence. For example, for the sequence $(a_n) = 1, 4, 9, 16, 25, 36, \ldots, n^2, \ldots$, the even subsequence (a_{2k}) is $4, 16, 36, \ldots, (2k)^2, \ldots$. Similarly the odd subsequence consists of the odd-numbered terms $a_1, a_3, a_5, \ldots$.

Problem 1.14

Show that (2^{k^2}) is a subsequence of (2^n).

The terms of (2^{k^2}) are $2, 2^4, 2^9, 2^{16}, \ldots$.

The following result relates the convergence of a given sequence to the convergence of its subsequences.

Theorem 1.3 Convergence of subsequences

If (a_n) is a convergent sequence with limit l, then every subsequence (a_{n_k}) is convergent with the same limit l.

We ask you to prove this theorem in the problems for this unit.

Remark

Theorem 1.3 can be used to show that a given sequence is *divergent*, since it tells us that:

- ► if (a_n) has a divergent subsequence, then (a_n) is itself divergent;
- ► if a sequence (a_n) has two convergent subsequences whose limits are different, then (a_n) is divergent.

Worked problem 1.8

Show that $(a_n) = ((-1)^n)$ is a divergent sequence.

In Worked problem 1.4 we proved that this sequence is not a null sequence.

Solution

We produce two subsequences of (a_n) that converge to different limits.

Consider the even subsequence $(a_{2k}) = ((-1)^{2k})$. Each term of this subsequence is 1, and so this subsequence converges to 1.

Now consider the odd subsequence $(a_{2k-1}) = ((-1)^{2k-1})$. Each term is -1, and so this subsequence converges to -1, a different limit.

By Theorem 1.3, we conclude that $(a_n) = ((-1)^n)$ is divergent. ■

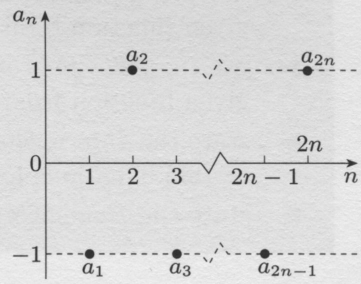


Figure 1.11

Problem 1.15

Determine whether the following sequence is convergent or divergent, justifying your answer.

$$\left((-1)^n \frac{1-n}{n}\right)$$

2 Continuity on the real line

After working through this section, you should be able to:

- use the *sequential definition of continuity* to determine whether a given function from $\mathbb{R}$ to $\mathbb{R}$ is continuous;
- use the *rules for continuous functions* to determine whether a given function from $\mathbb{R}$ to $\mathbb{R}$ is continuous;
- state the *Intermediate Value Theorem* and use it to show that certain equations have roots;
- state the *Boundedness* and *Extreme Value Theorems.*

In this section, we start our investigation of continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Initially, we use convergent sequences as our main tool for testing the continuity of functions. In Section 3, we give an alternative formulation of the definition of continuity that does not involve the use of convergent sequences.

2.1 Continuous and discontinuous functions

Consider the functions $f_1: \mathbb{R} \to \mathbb{R}$ and $f_2: \mathbb{R} \to \mathbb{R}$ given by

$$f_1(x) = x^2,$$

$$f_2(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

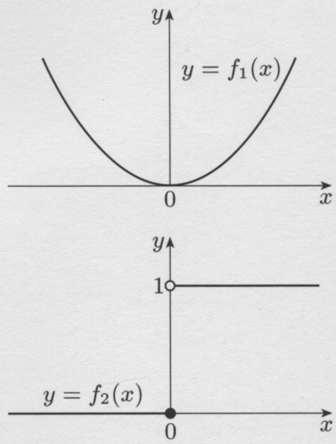


Figure 2.1

Intuitively, we would say that f_1 is continuous since its graph has no jumps or breaks in it, whereas f_2 is discontinuous since there is a jump at $x = 0$. We could go on to say that f_1 is continuous at every point $x \in \mathbb{R}$, whereas f_2 is discontinuous at $x = 0$.

Our immediate objective is to capture this intuitive idea of a continuous function in a precise way. We do this by using the idea that, if f is continuous at a point a and if x is a point near to a, then $f(x)$ should be near to $f(a)$. In other words, our definition should say in a precise way that

if x tends to a, then $f(x)$ tends to $f(a)$.

We make this notion exact by using the convergence of sequences.

Definition

Let $A \subseteq \mathbb{R}$ and let $f: A \to \mathbb{R}$ be a function.

Then f is **continuous** at $a \in A$ if:

for each sequence (x_n) in A for which $x_n \to a$ as $n \to \infty$,
the sequence $(f(x_n))$ converges to $f(a)$.

If f is not continuous at $a \in A$, then f is **discontinuous** at a.

Remarks

(i) This definition of continuity is often called the **sequential definition of continuity**, or simply **sequential continuity**, to distinguish it from the alternative definition that will be considered in Section 3.

(ii) To show that f is continuous at a we must show that $(f(x_n))$ converges to $f(a)$ for *every possible choice* of a sequence (x_n) that converges to a.

(iii) If the sequence $(f(x_n))$ converges to $f(a)$ when the sequence (x_n) converges to a, then we often write

$$f(x_n) \to f(a) \text{ as } x_n \to a.$$

If the sequence $(f(x_n))$ converges to $f(a)$ for *every* sequence (x_n) that converges to a, then we often write

$$f(x) \to f(a) \text{ as } x \to a.$$

Continuity as defined above is the property of a function *at a point*. It is easily extended to define continuity *on a set*.

Definition

Let $S \subseteq A \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ be a function.

We say that f is **continuous** on S when f is continuous at each point in S.

We say that f is **discontinuous** on S if f is discontinuous at at least one point in S.

Remarks

(i) If we simply say that a function $f\colon A \to \mathbb{R}$ is *continuous* (or *discontinuous*), by implication we mean that it is continuous (or discontinuous) on A.

(ii) To prove continuity on a set S, all that is needed is to prove it for a *general* point a in S.

(iii) To prove discontinuity on a set S, all that is needed is to prove it for a *single, specific* point in S.

We now show how the definition of continuity is used to prove that a function is continuous.

Worked problem 2.1

Prove that the function $f_1\colon \mathbb{R} \to \mathbb{R}$ given by

$$f_1(x) = x^2$$

is (a) continuous at 2 and (b) continuous on $\mathbb{R}$.

Solution

(a) By the definition of continuity we must show that, if (x_n) is any sequence in $\mathbb{R}$ that converges to 2, then the sequence $(f_1(x_n))$ converges to $f_1(2) = 4$.

Let (x_n) be any sequence that converges to 2. We know that $f_1(x_n) = x_n^2$. Therefore, by the Product Rule for sequences,

$$f_1(x_n) = x_n \times x_n \to 2 \times 2 = 4 \text{ as } n \to \infty.$$

But $f_1(2) = 4$, and so $f_1(x_n) \to f_1(2)$ as $x_n \to 2$.

Hence f_1 is continuous at 2.

(b) This is a generalization of the method in (a).

Let a be a general point in $\mathbb{R}$. By the definition of continuity we must show that, if (x_n) is any sequence in $\mathbb{R}$ that converges to a, then the sequence $(f_1(x_n))$ converges to $f_1(a) = a^2$.

Let (x_n) be any sequence that converges to a. By the Product Rule for sequences,

$$f_1(x_n) = x_n^2 = x_n \times x_n \to a \times a = a^2 \text{ as } n \to \infty.$$

But $f_1(a) = a^2$, and so $f_1(x_n) \to f_1(a)$ as $x_n \to a$.

Hence f_1 is continuous at a.

Since a is an arbitrary point of $\mathbb{R}$, f_1 is continuous on $\mathbb{R}$. ■

Problem 2.1

Prove that the function $f: \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = x^3 + 3$$

is continuous on $\mathbb{R}$.

To show that a function f is *discontinuous* at a point $a \in A$, we need to show that there is a sequence (x_n) in A for which

$$x_n \to a \text{ but } f(x_n) \not\to f(a).$$

Worked problem 2.2

Prove that the function $f_2: \mathbb{R} \to \mathbb{R}$ given by

$$f_2(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases}$$

is discontinuous at 0.

Solution

We must find a sequence (x_n) in $\mathbb{R}$ such that $x_n \to 0$ as $n \to \infty$, but $f_2(x_n) \not\to f_2(0)$ as $n \to \infty$.

Since $f_2(0) = 0$ and $f_2(x) = 1$ for $x > 0$, we choose any sequence (x_n) that converges to 0 'from the right', so that $f_2(x_n) = 1$ for all n. One such sequence is $(x_n) = (1/n)$. We have

$$x_n = 1/n \to 0 \text{ as } n \to \infty \text{ and } f_2(x_n) = 1 \text{ for all } n,$$
$$\text{so } f_2(x_n) \to 1 \text{ as } n \to \infty.$$

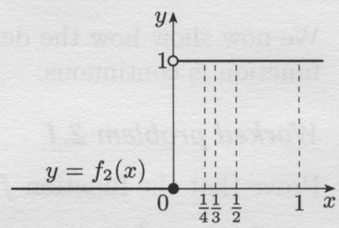


Figure 2.2

Since $f_2(0) = 0$ and $f_2(x_n) \not\to 0$,

$$f_2(x_n) \not\to f_2(0) \text{ as } n \to \infty.$$

It follows that f_2 is discontinuous at 0. ■

Problem 2.2

Prove that f_2 is continuous at a for each $a \neq 0$.

Thus f_2 is discontinuous only at $a = 0$. This is what we would expect from looking at its graph.

Now let us see what our definition tells us about the continuity of a function whose behaviour is less obvious.

Worked problem 2.3

Let $f\colon \mathbb{R} \to \mathbb{R}$ be the function defined by

$$f(x) = \begin{cases} \cos(\pi/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Decide whether f is continuous at 0, and justify your decision.

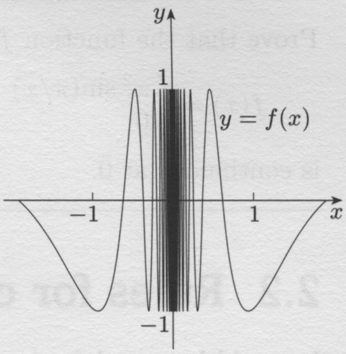


Figure 2.3 The graph of f for Worked problem 2.3

Solution

As Figure 2.3 illustrates, the graph oscillates very rapidly close to the origin, and there are points x, arbitrarily close to the origin, where $\cos(\pi/x)$ takes values away from 0. We guess that f is *discontinuous* at 0.

In order to prove this, we seek a sequence (x_n) in $\mathbb{R}$ for which

$$x_n \to 0 \text{ but } f(x_n) \not\to f(0) = 0.$$

We choose a sequence (x_n) of points for which $\cos(\pi/x_n) = 1$ (that is, the maximum points of the oscillation), so that $f(x_n) = \cos(\pi/x_n) = 1$ for all n and hence $f(x_n) \to 1 \neq 0$.

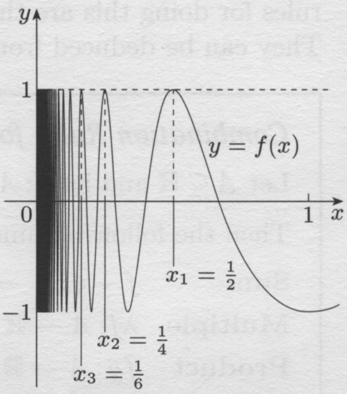


Figure 2.4 Choosing the sequence (x_n)

Now $\cos a = 1$ whenever a is of the form $2\pi n$, where n is an integer. It follows that $\cos(\pi/x) = 1$ whenever $\pi/x = 2\pi n$, i.e. whenever $x = 1/(2n)$. So if we take (x_n) to be the sequence given by $x_n = 1/(2n)$, then

$$f(x_n) = \cos\left(\frac{\pi}{x_n}\right) = \cos\left(\frac{\pi}{1/(2n)}\right) = \cos 2\pi n = 1 \quad \text{for all } n.$$

Moreover, $x_n \to 0$ as $n \to \infty$.

So we have a sequence (x_n) such that $x_n \to 0$ but $f(x_n) \not\to f(0)$.

Thus f is discontinuous at 0. ■

We now modify the function considered in Worked problem 2.3 to produce a function that *is* continuous at 0.

Worked problem 2.4

Prove that the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} x\cos(\pi/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is continuous at 0.

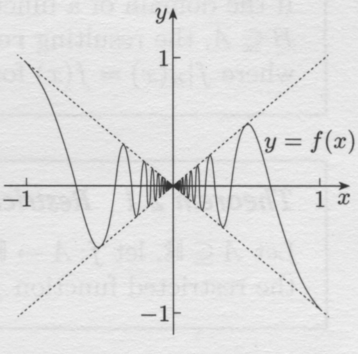


Figure 2.5 The graph of f for Worked problem 2.4

Solution

We need to show that, for any *null* sequence (x_n), $f(x_n) \to f(0) = 0$.

So suppose (x_n) is null. Then, for each n,

$$\begin{array}{lll} \textit{either} & x_n \neq 0 & \text{and} \quad |f(x_n)| = |x_n \cos(\pi/x_n)| = |x_n|\,|\cos(\pi/x_n)|, \\ \textit{or} & x_n = 0 & \text{and} \quad f(x_n) = 0. \end{array}$$

Now $|\cos a| \leq 1$ for all real numbers a and so, when $x_n \neq 0$,

$$|f(x_n)| = |x_n|\,|\cos(\pi/x_n)| \leq |x_n|.$$

When $x_n = 0$,

$$|f(x_n)| = 0 \leq |x_n| = 0.$$

Therefore $|f(x_n)| \leq |x_n|$ for all n.

But (x_n) is a null sequence and so, by the Squeeze Rule, $f(x_n) \to 0$ as $n \to \infty$. Thus, $f(x_n) \to f(0) = 0$ as $x_n \to 0$.

Hence f is continuous at 0. ■

Problem 2.3

Prove that the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} x^2 \sin(\pi/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is continuous at 0.

2.2 Rules for continuous functions

It would be very laborious to have to use the definition of continuity every time we wish to prove that a function is continuous. Fortunately, we can short-cut that process by using functions we already know to be continuous to deduce that other functions are continuous. Some of the most useful rules for doing this are the Combination Rules for continuous functions. They can be deduced from the Combination Rules for sequences.

Combination Rules for continuous functions from $\mathbb{R}$ to $\mathbb{R}$

Let $A \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ and $g\colon A \to \mathbb{R}$ be continuous on A.

Then the following functions are continuous on A:

Sum $f + g\colon A \to \mathbb{R}$, defined by $(f+g)(x) = f(x) + g(x)$;

Multiple $\lambda f\colon A \to \mathbb{R}$ where $\lambda \in \mathbb{R}$, defined by $(\lambda f)(x) = \lambda \times f(x)$;

Product $fg\colon A \to \mathbb{R}$, defined by $(fg)(x) = f(x)g(x)$.

Another useful result tells us that restricting the domain of a continuous function to a smaller set still gives us a continuous function. Before we state this result, we need a definition.

Definition

If the domain of a function $f\colon A \to \mathbb{R}$ is restricted to a set B, where $B \subseteq A$, the resulting **restricted function** is denoted by $f|_B : B \to \mathbb{R}$, where $f|_B(x) = f(x)$ for all $x \in B$.

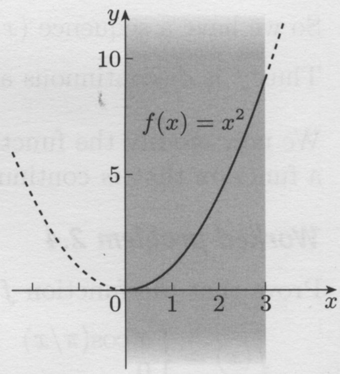


Figure 2.6 The function $f(x) = x^2$ restricted to $[0, 3]$

Theorem 2.1 Restriction Rule

Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$ be continuous on A and let $B \subseteq A$. Then the restricted function $f|_B : B \to \mathbb{R}$ is continuous on B.

A proof of this result is given in Section 3.

For example, the function $f(x) = x^2$ is continuous on $\mathbb{R}$. Let $B = [0, 3]$. Then $f|_B : B \to \mathbb{R}$ is f restricted to the interval $[0, 3]$ and is continuous on $[0, 3]$, as Figure 2.6 illustrates.

Our final useful result about combining continuous functions allows us to compose continuous functions to form a continuous function. Before stating this result, we remind you of the definition of the *composition* of two functions.

> ### *Definition*
>
> If $f\colon A \to C$ and $g\colon B \to D$ are functions with $f(A) \subseteq B$, then the **composed function** or **composite** $g \circ f\colon A \to D$ is given by $(g \circ f)(x) = g(f(x))$.

$$
\begin{tikzcd}
 & f(A) \subseteq B \arrow[dr, "g"] & \\
A \arrow[ur, "f"] \arrow[rr, bend right, "g \circ f"'] & & D
\end{tikzcd}
$$

Figure 2.7

Remark

The condition $f(A) \subseteq B$ is needed so that the image set of the function f lies inside the domain of g. Some definitions of composition omit this condition, but we shall require it to hold throughout this course.

For example, let $f\colon \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^3$ and let $g\colon \mathbb{R} \to [-1, 1]$ be given by $g(x) = \sin x$. Then $g \circ f\colon \mathbb{R} \to [-1, 1]$ is given by $(g \circ f)(x) = \sin(x^3)$.

Here $A = B = C = \mathbb{R}$, $D = [-1, 1]$ and $f(A) = \mathbb{R} \subseteq B$.

> ### *Theorem 2.2 Composition Rule*
>
> Let A, $B \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ and $g\colon B \to \mathbb{R}$ be continuous on A and B respectively, with $f(A) \subseteq B$. Then the composed function $g \circ f\colon A \to \mathbb{R}$ is continuous on A.

A proof of this result is given in Section 3.

These rules enable us to use a collection of basic continuous functions to form new functions that must then also be continuous.

> ### *Basic continuous functions*
>
> The following functions are continuous:
>
> - polynomials and rational functions;
> - $f(x) = |x|$;
> - $f(x) = \sqrt{x}$;
> - the trigonometric functions (sine, cosine and tangent) and their inverses;
> - the logarithmic function~~s~~;
> - the exponential function~~s~~.

Note that the domains of these basic continuous functions exclude points at which the functions are undefined. For example, the domain of a rational function excludes any points where the denominator is zero, the domain of $f(x) = \sqrt{x}$ is $[0, \infty)$, and the domain of tan excludes the points $(2n + 1)\pi/2$ for all $n \in \mathbb{Z}$.

The application of the rules to functions constructed from basic continuous functions is illustrated in the following worked problem.

Worked problem 2.5

Using the above rules for continuous functions, show that the function $f\colon [-\pi, \pi] \to \mathbb{R}$ given by

$$f(x) = e^x + 3\cos 2x$$

is continuous on $[-\pi, \pi]$.

Solution

The functions $x \mapsto 2x$, $x \mapsto e^x$ and $x \mapsto \cos x$ are all basic continuous functions, defined on the whole of $\mathbb{R}$.

By the Restriction Rule, $x \mapsto e^x$ and $x \mapsto 2x$ are both continuous functions on the interval $[-\pi, \pi]$.

The image set of $x \mapsto 2x$ restricted to $[-\pi, \pi]$ is $[-2\pi, 2\pi] \subseteq \mathbb{R}$. Thus, by the Composition Rule, $x \mapsto \cos 2x$ is continuous on $[-\pi, \pi]$. So, by the Multiple Rule, $x \mapsto 3\cos 2x$ is continuous on $[-\pi, \pi]$.

Finally, by the Sum Rule, f is continuous on $[-\pi, \pi]$. ■

Problem 2.4

Show that $f\colon (0, \infty) \to \mathbb{R}$ given by

$$f(x) = 2\sin(1/x) - 3e^{-4x^3}$$

is continuous on $(0, \infty)$.

Problem 2.5

Let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuous function. Show that the function $|f|\colon \mathbb{R} \to \mathbb{R}$ given by $|f|(x) = |f(x)|$ is continuous on $\mathbb{R}$.

2.3 Continuity theorems

In this subsection we discuss three important theorems concerning continuous functions from $\mathbb{R}$ to $\mathbb{R}$. These theorems, which we shall not prove, help us to investigate properties of continuous functions from $\mathbb{R}$ to $\mathbb{R}$. When we extend the notion of continuity to more general settings, one of our concerns will be to see how far these theorems still hold true.

You may well have seen proofs of these theorems elsewhere. They are given in most courses on real analysis.

All three theorems concern continuous functions whose domain is a *closed bounded interval*, so we need to be clear what this means. Recall that a **closed interval** is one that includes its endpoints. It can take one of the following forms:

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\};$$
$$[a, \infty) = \{x \in \mathbb{R} : x \geq a\};$$
$$(-\infty, b] = \{x \in \mathbb{R} : x \leq b\};$$
$$(-\infty, \infty) = \mathbb{R}.$$

An *open interval*, by contrast, does not include its endpoints. Open intervals are of the forms:

$$(a, b) = \{x \in \mathbb{R} : a < x < b\};$$
$$(a, \infty) = \{x \in \mathbb{R} : x > a\};$$
$$(-\infty, b) = \{x \in \mathbb{R} : x < b\};$$
$$(-\infty, \infty) = \mathbb{R}.$$

Note that $(-\infty, \infty)$ is both closed and open.

For an interval to be *bounded* there must be some real number M such that $|x| \leq M$ for all x in the interval. Thus only the first type of closed interval above is bounded. So the **closed bounded intervals** are just those of the form $[a, b]$.

The Intermediate Value Theorem

Our first theorem says that continuous functions do not 'omit any values', and so there cannot be any jumps or gaps in their graphs. Thus, when f is continuous on a closed interval and attains two particular values, it must also take all the values between them. For example, let f be continuous on $[a, b]$, and suppose that $f(a) < 0$ and $f(b) > 0$. Then the theorem asserts that there must be at least one point $c \in (a, b)$ for which $f(c) = 0$ — that is, the graph of f must cross the x-axis, as Figure 2.8 illustrates.

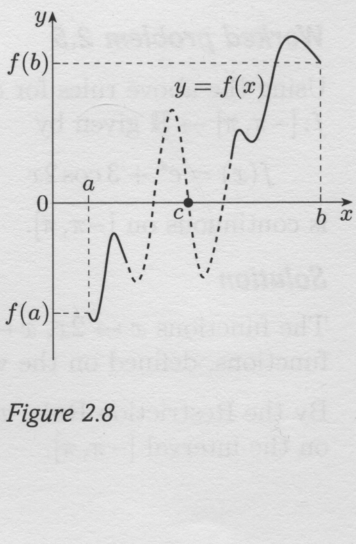


Figure 2.8

> ***Theorem 2.3 Intermediate Value Theorem***
>
> Let $f\colon [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and let k be any number between $f(a)$ and $f(b)$. Then there exists a number $c \in (a, b)$ such that $f(c) = k$.

Remark

This theorem can be adapted to apply to all intervals and not just to closed bounded intervals. However, it does not hold if the domain of f is not an interval. For example, consider the function $f(x) = 1/x$ restricted to the set $[-1, 1] - \{0\}$, as shown in Figure 2.9. This function is continuous on $[-1, 1] - \{0\}$ and has $f(-1) < 0$ and $f(1) > 0$, but there is no value of $x \in [-1, 1] - \{0\}$ for which $f(x) = 0$.

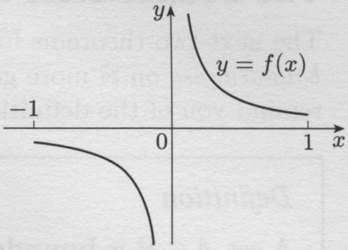


Figure 2.9

Worked problem 2.6

Let $f\colon [-\pi, \pi] \to \mathbb{R}$ be the function given by

$$f(x) = e^x + 3\cos 2x.$$

Prove that there is at least one point $c \in [-\pi, \pi]$ for which $f(c) = 0$.

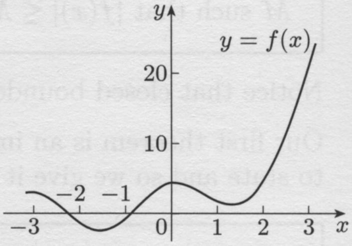


Figure 2.10 The graph of $f(x) = e^x + 3\cos 2x$ *on* $[-\pi, \pi]$

Solution

We saw in Worked problem 2.5 that f is continuous on $[-\pi, \pi]$.

To show that there is a point $c \in [-\pi, \pi]$ for which $f(c) = 0$, using the Intermediate Value Theorem, we need an interval $[a, b]$ for which $f(a) < 0$ and $f(b) > 0$. We cannot use the endpoints of $[-\pi, \pi]$ because both $f(-\pi)$ and $f(\pi)$ are positive. So we need to restrict our attention to a suitable subinterval of $[-\pi, \pi]$. The set $[-\pi/2, 0]$ is a closed interval contained in $[-\pi, \pi]$, and the Restriction Rule implies that f is continuous on it. Moreover,

$$f(-\pi/2) = e^{-\pi/2} + 3\cos(-\pi) = e^{-\pi/2} - 3 < 0,$$

$$f(0) = e^0 + 3\cos 0 = 1 + 3 = 4 > 0.$$

$e^{-\pi/2} < e^0 = 1.$

Since the conditions of the Intermediate Value Theorem are satisfied by f on $[-\pi/2, 0]$, we deduce that there is a number c (with $-\pi/2 < c < 0$) for which $f(c) = 0$. ■

The Intermediate Value Theorem is an *existence result*: it tells us that there is a point in the interval whose image under f is 0, but it does not tell us how to find such a point. Thus, although we can deduce that the function $f\colon [-\pi, \pi] \to \mathbb{R}$ given by $f(x) = e^x + 3\cos 2x$ takes the value 0 at some point(s) in the interval $(-\pi/2, 0)$, we must use different methods to find the value(s) of x for which this is true.

The unique solution to $e^x + 3\cos 2x = 0$ in $[-\pi/2, 0]$ is approximately $x = -0.8564$.

Although the theorem does not provide a method for finding a point c at which a function f takes an intermediate value k, it is nevertheless very useful. It enables us to know in advance whether such a point c exists, and it applies to *all* continuous functions defined on a closed bounded interval.

Problem 2.6

The function $f\colon \mathbb{R} \to \mathbb{R}$ is given by

$$f(x) = \log(1 + x^2) + 3\cos \pi x.$$

Prove that there is a number c in the open interval $(0, 3)$ for which $f(c) = 0$.

Problem 2.7 *A fixed-point theorem*

Let $f\colon [0, 1] \to [0, 1]$ be a continuous function. Use the Intermediate Value Theorem to show that there is a number $k \in [0, 1]$ such that $f(k) = k$.

Hint Consider the function $g\colon [0, 1] \to \mathbb{R}$ defined by $g(x) = f(x) - x$.

Problem 2.7 shows that each continuous function $f\colon [0, 1] \to [0, 1]$ has a *fixed point*, a point that maps to itself under f.

The Boundedness and Extreme Value Theorems

The next two theorems involve not only closed bounded intervals but *boundedness* on $\mathbb{R}$ more generally. So, before we discuss the theorems, we remind you of the definition of boundedness on $\mathbb{R}$.

> ***Definition***
>
> A set $A \subset \mathbb{R}$ is **bounded** if there exists a real number M such that $|x| \leq M$ for all $x \in A$.
>
> A function $f: A \to \mathbb{R}$ is **bounded** on A if there exists a real number M such that $|f(x)| \leq M$ for all $x \in A$.

Notice that closed bounded intervals $[a, b]$ are bounded sets.

Open intervals of the form (a, b) are also bounded sets.

Our first theorem is an immediate consequence of the second, but is easier to state and so we give it first.

> ***Theorem 2.4 Boundedness Theorem***
>
> Let $f: [a, b] \to \mathbb{R}$ be a continuous function on $[a, b]$. Then f is bounded on $[a, b]$.

Remark

This theorem states that continuous functions on closed bounded intervals must be bounded: that is, the image set of the closed bounded interval $[a, b]$ under f is a bounded set. Note that this is not necessarily the case if the domain is not bounded: for example, the function $f: \mathbb{R} \to \mathbb{R}$ given by $f(x) = x$ is a continuous function, and the image set of $\mathbb{R}$ under f is again $\mathbb{R}$, an *un*bounded set.

Our next result tells us that a continuous function on a closed bounded interval is not only bounded but attains its maximum and minimum values.

> ***Theorem 2.5 Extreme Value Theorem***
>
> Let $f: [a, b] \to \mathbb{R}$ be a continuous function on $[a, b]$. Then there exist numbers $c, d \in [a, b]$ such that
>
> $$f(c) \leq f(x) \leq f(d) \quad \text{for all } x \in [a, b].$$

Remarks

(i) The theorem does not hold if the domain of f is not both closed and bounded.

(ii) The number c is such that $f(c) \leq f(x)$, for all $x \in [a, b]$ — that is, $f(c)$ is the *minimum* value of the function in the interval. Similarly, $f(d)$ is the *maximum* value of the function in the interval.

Example 2.1

Consider $f\colon [-\pi, \pi] \to \mathbb{R}$ defined by $f(x) = e^x + 3\cos 2x$. Because f is continuous on $[-\pi, \pi]$ (see Worked problem 2.5), we can use the Extreme Value Theorem to deduce that there are points c and d in $[-\pi, \pi]$ such that $f(c) \le f(x) \le f(d)$ for all $x \in [-\pi, \pi]$. As Figure 2.11 shows, c lies near -1.6 while $d = \pi$. ■

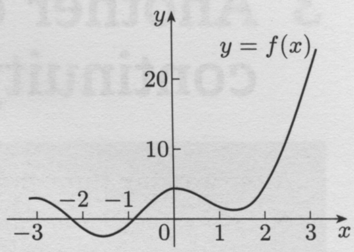


Figure 2.11 The graph of $f(x) = e^x + 3\cos 2x$ *on* $[-\pi, \pi]$

The Extreme Value Theorem is another example of an existence result: again, it does not tell us how to find the points whose existence it guarantees — we have to use other methods to find them.

Note that the Intermediate Value, Boundedness and Extreme Value Theorems hold for *all continuous functions* on *all closed bounded intervals.* We conclude that they are the result of an interaction between continuity and closed bounded intervals. When we look at the properties of more general spaces in later units, a generalization of the idea of closed bounded intervals gives rise to the notion of 'compactness'.

3 Another definition of continuity

After working through this section, you should be able to:

- state the ε–δ *definition of continuity*;
- use the ε–δ definition of continuity to determine whether a given function from $\mathbb{R}$ to $\mathbb{R}$ is continuous;
- appreciate the equivalence of the sequential and ε–δ definitions of continuity.

Our earlier definition of continuity in terms of sequences is perfectly adequate for working with functions from $\mathbb{R}$ to $\mathbb{R}$; indeed a similar definition can also be framed for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, for any natural numbers n and m.

A central feature of that definition is the use of the convergence of sequences to pin down the idea of 'nearness' to a point a. There is another definition of continuity which describes nearness in terms of open intervals around a. This new definition, involving ideas of 'distance', is explored in this section.

Why do we need it? One reason is that this new definition of continuity generalizes more naturally to the kinds of spaces we shall be dealing with later. An additional benefit is that it is more suited than the sequential definition to showing the continuity of certain useful functions.

Since this new definition of continuity is the one upon which all our later work depends, it is important that you understand and become familiar with it. To this end, there is some software to accompany this section, which we suggest you work through after you have studied Subsection 3.2. Most of this section will be spent developing the consequences of this new definition of continuity. At the end of this section we show that the two definitions of continuity are equivalent.

3.1 The ε–δ definition of continuity

Any definition of continuity needs to say in a precise way that a function f is continuous at a point a if $f(x)$ is close to $f(a)$ whenever x is close to a. Before stating the new definition we illustrate by means of an example how this can be done.

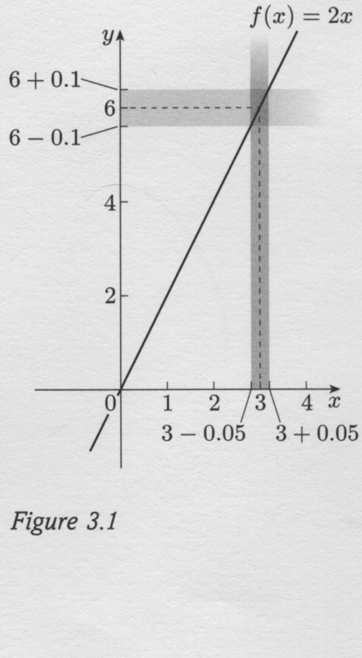


Figure 3.1

Example 3.1

Consider the function $f(x) = 2x$ at the point 3. If x is any number close to 3, then $2x$ is close to 6 — that is, $f(x)$ is close to $f(3)$. In fact, we can arrange that $f(x)$ is as close as we like to $f(3)$ by making x close enough to 3. For example, we can arrange that $f(x)$ is within 0.1 of 6 by requiring x to be within 0.05 of 3, as Figure 3.1 illustrates. In symbols,

$$6 - 0.1 < f(x) < 6 + 0.1 \quad \text{whenever} \quad 3 - 0.05 < x < 3 + 0.05.$$

Using modulus symbols we can write this as

$$|f(x) - 6| < 0.1 \quad \text{whenever} \quad |x - 3| < 0.05.$$

Similarly, we can arrange that $f(x)$ is within 0.01 of 6 by requiring x to be within 0.005 of 3. In symbols,

$$6 - 0.01 < f(x) < 6 + 0.01 \quad \text{whenever} \quad 3 - 0.005 < x < 3 + 0.005,$$

or equivalently, using modulus symbols,

$$|f(x) - 6| < 0.01 \quad \text{whenever} \quad |x - 3| < 0.005.$$

In fact, whatever positive number ε we choose, we can find a corresponding positive number δ such that

$$|f(x) - 6| < \varepsilon \quad \text{whenever} \quad |x - 3| < \delta$$

simply by taking $\delta = \varepsilon/2$. ■

We now give the definition of continuity at a point a.

Definition

Let $A \subseteq \mathbb{R}$. A function $f\colon A \to \mathbb{R}$ is **continuous** at $a \in A$ if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that, for all $x \in A$,

$$|f(x) - f(a)| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta.$$

Remarks

(i) We refer to this as the **ε–δ definition of continuity**, or simply as **ε–δ continuity**, when we need to distinguish it from the sequential definition of continuity, given in Section 2.

(ii) It is important to be able to interpret the concise statement

$$|f(x) - f(a)| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta.$$

As we have seen, $|f(x) - f(a)|$ denotes the *distance* between $f(x)$ and $f(a)$ and $|x - a|$ denotes the *distance* between x and a. So the statement can be interpreted as:

the distance between $f(x)$ and $f(a)$ is less than ε whenever the distance between x and a is less than δ.

Since $|f(x) - f(a)| < \varepsilon$ means $f(a) - \varepsilon < f(x) < f(a) + \varepsilon$ and $|x - a| < \delta$ means $a - \delta < x < a + \delta$, the statement can also be written as

$$f(a) - \varepsilon < f(x) < f(a) + \varepsilon \quad \text{whenever} \quad a - \delta < x < a + \delta.$$

This says that $f(x)$ is within an open interval of width 2ε centred at $f(a)$ whenever x is within the open interval of width 2δ centred at a, as Figure 3.2 illustrates.

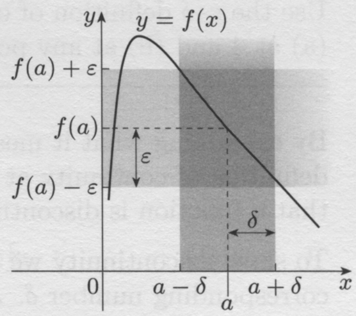


Figure 3.2

Worked problem 3.1

Let $f\colon \mathbb{R} \to \mathbb{R}$ be given by

$$f(x) = 3x + 5.$$

Use the ε–δ definition of continuity to prove that f is continuous at any point $a \in \mathbb{R}$.

Solution

Fix $a \in \mathbb{R}$. We must show that, for each positive number ε, there is a positive number δ such that

$$|f(x) - f(a)| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta.$$

Let $\varepsilon > 0$ be given.

First, we express $|f(x) - f(a)|$ in terms of $|x - a|$:

$$|f(x) - f(a)| = |(3x + 5) - (3a + 5)| = |3x - 3a| = |3(x - a)| = 3|x - a|.$$

So, when $|x - a| < \delta$,

$$|f(x) - f(a)| = 3|x - a| < 3\delta.$$

Thus, by requiring that the distance between x and a is less than δ, we can ensure that the distance between $f(x)$ and $f(a)$ is less than 3δ. Hence, if we take $\delta = \varepsilon/3$, so that $\varepsilon = 3\delta$, then $|f(x) - f(a)| < \varepsilon$ whenever $|x - a| < \delta$.

We deduce that f is continuous at a. ■

The procedure used here was:

- ▶ express $|f(x) - f(a)|$ in terms of $|x - a|$;
- ▶ use this information to obtain a relationship between ε and δ which can be used to produce an expression for δ.

This is the general procedure for showing continuity using the ε–δ definition.

Problem 3.1

Let $f: \mathbb{R} \to \mathbb{R}$ be given by

$$f(x) = 3 - 2x.$$

Use the ε–δ definition of continuity to prove that $f(x)$ is continuous (a) at 4 and (b) at any point $a \in \mathbb{R}$.

By examining what it means when a function fails to satisfy our new definition of continuity at a point, we find an alternative way of saying that a function is discontinuous.

To show discontinuity we need to find a number ε for which there is no corresponding number δ. *Just one ε that 'goes wrong' will do.* Thus, with this value of ε, for *all* positive values of δ we can find x with $|x - a| < \delta$ such that $|f(x) - f(a)| \geq \varepsilon$.

> ***Definition***
>
> Let $A \subseteq \mathbb{R}$. The function $f: A \to \mathbb{R}$ is **discontinuous** at $a \in A$ if it is not continuous at a. That is, if there is an $\varepsilon > 0$ such that, for *all* $\delta > 0$, there is an $x \in A$ with both
>
> $$|x - a| < \delta \quad \text{and} \quad |f(x) - f(a)| \geq \varepsilon.$$

Remarks

(i) This definition is just the negation of the definition of continuity. Thus a function is discontinuous at a if and only if it is not continuous at a.

(ii) It is worth comparing the conditions in this definition — 'there is an $\varepsilon > 0$', 'for all $\delta > 0$', 'there is an $x \in A$' — with those for the definition for continuity. For example, 'for each $\varepsilon > 0$' now becomes 'there is an $\varepsilon > 0$'. For continuity, the condition is for *every* ε, whereas for discontinuity the condition is that there is (at least) *one* ε. The other conditions are changed similarly.

Informally, this definition says: a function is discontinuous at a point a when, whatever δ-interval around a is chosen, it produces an interval about $f(a)$ whose half-width is always greater than or equal to some fixed value of ε.

For example, consider the function $f_2 \colon \mathbb{R} \to \mathbb{R}$, which was defined in Section 2 by

$$f_2(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

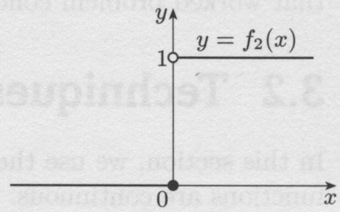


Figure 3.3

Take $\varepsilon = \frac{1}{2}$. Can you find any δ-interval about 0 such that, for all x in that interval, $|f_2(x) - f_2(0)| < \frac{1}{2}$? There is no such δ-interval. Whatever δ-interval you consider, it always contains some values of $f_2(x)$ equal to 1 and some values equal to 0. So, for some x in that δ-interval, $|f_2(x) - f_2(0)| \geq \varepsilon$. The discontinuity of f_2 at 0 is proved formally in the following worked problem.

Worked problem 3.2

Use the ε–δ definition of continuity to prove that the function $f_2 \colon \mathbb{R} \to \mathbb{R}$ given by

$$f_2(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases}$$

is discontinuous at 0.

Solution

To show that f_2 is discontinuous at 0, it is enough to find a single $\varepsilon > 0$ such that, for each $\delta > 0$, we can find an $x \in \mathbb{R}$ with both

You may like to compare this solution with the solution to Worked problem 2.2.

$$|x - 0| < \delta \quad \text{and} \quad |f_2(x) - f_2(0)| \geq \varepsilon.$$

Since $f_2(0) = 0$, we need to find an x such that $|x| < \delta$ and $|f_2(x)| \geq \varepsilon$. If we set $\varepsilon = \frac{1}{2}$ and take $x = \delta/2$, then

$$|x| = \delta/2 < \delta \quad \text{and} \quad |f_2(x)| = 1 \geq \tfrac{1}{2} = \varepsilon.$$

We conclude that f_2 is discontinuous at 0. ■

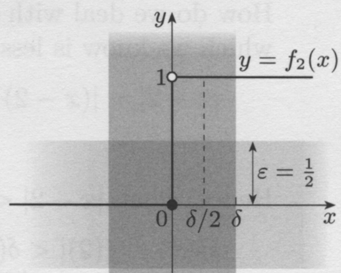


Figure 3.4

We have been considering the continuity of a function at individual points in the domain. However, it is only a short step to define continuity on a set. The definition is the same as in Section 2.

> ### *Definition*
>
> Let $S \subseteq A \subseteq \mathbb{R}$ and let $f \colon A \to \mathbb{R}$ be a function.
>
> We say that f is **continuous** on S when f is continuous at each point in S.
>
> We say that f is **discontinuous** on S if f is discontinuous at at least one point in S.

Remarks

(i) If we simply say that a function $f: A \to \mathbb{R}$ is *continuous* (or *discontinuous*), by implication we mean that it is continuous (or discontinuous) on A.

(ii) To prove continuity on a set S, all that is needed is to prove it for a *general* point a in S.

(iii) To prove discontinuity on a set S, all that is needed is to prove it for a *single, specific* point in S.

In light of this definition, we can deduce from Worked problem 3.1 that $f: \mathbb{R} \to \mathbb{R}$ given by $f(x) = 3x + 5$ is continuous on $\mathbb{R}$, since the proof in that worked problem concerned an arbitrary point $a \in \mathbb{R}$.

3.2 Techniques for proving continuity

In this section, we use the ε–δ definition of continuity to show that various functions are continuous.

Worked problem 3.3

Prove that the function $f: \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = x^2$$

is (a) continuous at 2 and (b) continuous on $\mathbb{R}$.

Solution

(a) Let $\varepsilon > 0$ be given.

We have to find $\delta > 0$ such that

$$|f(x) - f(2)| < \varepsilon \quad \text{whenever} \quad |x - 2| < \delta.$$

Now

$$|f(x) - f(2)| = |x^2 - 4| = |(x-2)(x+2)| = |x-2|\,|x+2|.$$

So, when $|x - 2| < \delta$,

$$|f(x) - f(2)| = |x-2|\,|x+2| < \delta|x+2|.$$

How do we deal with $|x + 2|$? Can we write it in terms of $|x - 2|$, which we know is less than δ? In fact we can, as follows:

$$\begin{aligned} |x+2| = |(x-2)+4| &\le |x-2| + 4 \qquad \text{(by the Triangle Inequality)} \\ &< \delta + 4 \qquad \text{(since } |x-2| < \delta\text{)}. \end{aligned}$$

Hence, when $|x - 2| < \delta$,

$$|f(x) - f(2)| < \delta(\delta + 4).$$

We can thus ensure that $|f(x) - f(2)| < \varepsilon$ by finding (in terms of ε) the value of δ that solves the equation $\delta(\delta + 4) = \varepsilon$. However, this leads to a rather complicated expression for δ and, after all, we only need to find a value of δ that 'works' — we do not have to find the best possible choice. In particular, we can assume that $\delta \le 1$. Thus

Solving this equation gives $\delta = -2 + \sqrt{4 + \varepsilon}$.

$$\delta(\delta + 4) \le \delta(1 + 4) = 5\delta.$$

We can now make $\delta(\delta + 4) \le \varepsilon$ by taking $\delta \le \varepsilon/5$. Hence, if $\delta = \min\{1, \varepsilon/5\}$, then $|f(x) - f(2)| < \varepsilon$ whenever $|x - 2| < \delta$.

Recall that $\min\{1, \varepsilon/5\}$ denotes the lesser of 1 and $\varepsilon/5$.

We conclude that f is continuous at 2.

(b) To prove that f is continuous on $\mathbb{R}$, we follow the same procedure but use a general point a.

Let $\varepsilon > 0$ be given.

We have to find $\delta > 0$ such that

$$|f(x) - f(a)| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta.$$

Now

$$|f(x) - f(a)| = |x^2 - a^2| = |(x-a)(x+a)| = |x-a|\,|x+a|.$$

So, when $|x - a| < \delta$,

$$|f(x) - f(a)| = |x-a|\,|x+a| < \delta|x+a|.$$

Using the Triangle Inequality, we have

$$|x+a| = |(x-a) + 2a| \leq |x-a| + |2a| < \delta + 2|a|.$$

Hence, when $|x - a| < \delta$,

$$|f(x) - f(a)| < \delta(\delta + 2|a|).$$

Assume that $\delta \leq 1$; then

$$\delta(\delta + 2|a|) \leq \delta(1 + 2|a|).$$

We can now make $\delta(\delta + 2|a|) \leq \varepsilon$ by taking $\delta \leq \dfrac{\varepsilon}{1 + 2|a|}$.

Hence, if $\delta = \min\left\{1, \dfrac{\varepsilon}{1+2|a|}\right\}$, then $|f(x) - f(a)| < \varepsilon$ whenever $|x - a| < \delta$.

We conclude f is continuous at a.

Since a is an arbitrary point of $\mathbb{R}$, f is continuous on $\mathbb{R}$. ■

Remarks

(i) In this solution we employed two very useful strategies.

The first was to rewrite $|x+2|$ as $|(x-2)+4|$ and use the Triangle Inequality.

The second was to assume that δ is no greater than some fixed positive number (in this case, 1). This means that the result holds only for values of $\delta \leq 1$, but as we can take δ to be *any* positive value that 'works' for the given ε, this is acceptable. This strategy is often useful when we wish to produce a simple expression for δ in terms of ε.

(ii) In (b), we arrived at the value $\delta \leq \dfrac{\varepsilon}{1+2|a|}$. Notice that here δ depends not only on ε, but also on a. This means that a different value of δ is needed for each point a.

Problem 3.2

Prove that the function $f: \mathbb{R} \to \mathbb{R}$, given by

$$f(x) = 3x^2 + 2$$

is (a) continuous at 4 and (b) continuous on $\mathbb{R}$.

Worked problem 3.4

Prove that the function $f: \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} \sqrt{x} & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases}$$

is continuous at a for each $a > 0$.

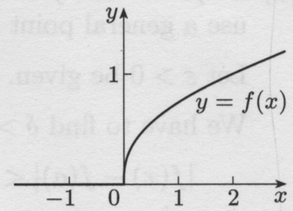


Figure 3.5 The function of Worked problem 3.4

Remark

The function f can be viewed as two different functions stitched together at 0, and the proof of the continuity of f on $\mathbb{R}$ can be split into three cases: $a > 0$, $a = 0$ and $a < 0$. This device of splitting the domain into regions where the behaviour of f is different avoids some of the complications that arise when considering continuity at a general point of $\mathbb{R}$.

In Problem 3.3 below, we ask you to prove that f is continuous when $a = 0$. You are asked to consider the case $a < 0$ in the problems for this unit.

Solution

Let $a > 0$, so that $f(a) = \sqrt{a}$. We must show that, for each positive number ε, there is a positive number δ such that

$$|f(x) - \sqrt{a}| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta.$$

Let $\varepsilon > 0$ be given. Consider those $x \in \mathbb{R}$ for which $|x - a| < \delta$, where δ is to be determined.

We wish to find an upper bound for $|f(x) - f(a)|$ in terms of δ. But there is a problem: $f(x)$ can be either 0 or $\sqrt{x}$ depending on whether $x > 0$ or $x \leq 0$. It would be helpful if we could ensure that x is greater than 0 so that $f(x) = \sqrt{x}$. Since $|x - a| < \delta$ is equivalent to saying that $a - \delta < x < a + \delta$, we see that if we choose δ so that $a - \delta \geq 0$, then x is necessarily positive and hence $f(x) = \sqrt{x}$. Since $a > 0$, we can always choose such a δ. Thus we assume that $\delta \leq a$, so that $x > 0$.

Remember, we need only to find a $\delta > 0$ that 'works' for ε. We can always assume that δ is less than some fixed positive number.

Now

$$|f(x) - f(a)| = |\sqrt{x} - \sqrt{a}|.$$

This expression is awkward to work with directly, so we write

$$\begin{aligned} |f(x) - f(a)| &= |\sqrt{x} - \sqrt{a}| \times \frac{\sqrt{x} + \sqrt{a}}{\sqrt{x} + \sqrt{a}} \\ &= \frac{|(\sqrt{x} - \sqrt{a})(\sqrt{x} + \sqrt{a})|}{\sqrt{x} + \sqrt{a}} = \frac{|x - a|}{\sqrt{x} + \sqrt{a}}. \end{aligned}$$

$(\sqrt{x} - \sqrt{a})(\sqrt{x} + \sqrt{a}) = x - a.$

Since $x > 0$, we have $\sqrt{x} + \sqrt{a} > \sqrt{a}$. Hence

$$|f(x) - f(a)| < \frac{|x - a|}{\sqrt{a}}.$$

We are assuming that $|x - a| < \delta$, and so

$$|f(x) - f(a)| < \frac{\delta}{\sqrt{a}}.$$

Thus, if we also choose δ to ensure that $\delta/\sqrt{a} \leq \varepsilon$, then we have $|f(x) - f(a)| < \varepsilon$. This is achieved by choosing $\delta \leq \varepsilon\sqrt{a}$. Hence, if $\delta = \min\{a, \varepsilon\sqrt{a}\}$, then $|f(x) - f(a)| < \varepsilon$ whenever $|x - a| < \delta$.

Thus f is continuous at a. ■

Remarks

(i) Notice that, once again, the choice for δ depends on the values of both ε and a.

(ii) It is usual to have to manipulate the expression for $|f(x) - f(a)|$ into a suitable form before you can apply an inequality involving δ.

Problem 3.3

Prove that the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} \sqrt{x} & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases}$$

is continuous at 0.

Hint Notice that, for all $x \in \mathbb{R}$, $|f(x)| \leq \sqrt{|x|}$.

The basic method that we use to show that a function is continuous, using the ε–δ definition, can be summarized as follows.

Strategy for using the ε–δ definition of continuity

▶ *To show that f is continuous at $a \in A$*

(a) Let $\varepsilon > 0$ be given.

(b) If the domain is split, try to choose $\delta > 0$ small enough to guarantee that, if $|x - a| < \delta$, then x lies in the same part of the domain of f as a.

(c) Suppose that $|x - a| < \delta$, and use the rule for f to find an upper bound for $|f(x) - f(a)|$ in terms of δ.

(d) Use this upper bound to choose δ so that, for all $x \in A$,

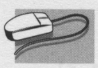
$$|f(x) - f(a)| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta.$$

▶ *To show that f is discontinuous at $a \in A$*

Find **one** positive number ε such that, for **each** positive number δ, there exists an $x \in A$ satisfying

$$|x - a| < \delta \quad \text{and} \quad |f(x) - f(a)| \geq \varepsilon.$$

Further insight into the ε–δ definition of continuity and its use can be obtained by using the software associated with this unit. We suggest that you work through this software now.



3.3 Rules for continuous functions

In Subsection 2.2 we stated several rules for combining continuous functions. These rules — the Combination, Restriction and Composition Rules — enable us to avoid working directly from the definition of continuity when we want to prove that a given function is continuous. Although these rules were introduced when we were using the sequential definition of continuity, the same rules hold for the ε–δ definition. In this subsection we prove the Restriction Rule and the Composition Rule using the ε–δ definition of continuity and give examples of their use.

We first restate the Restriction Rule and give a proof using the ε–δ definition of continuity.

> ***Theorem 3.1 Restriction Rule***
>
> Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$ be continuous on A and let $B \subseteq A$. Then the restricted function $f|_B\colon B \to \mathbb{R}$ is continuous on B.

Proof Let $a \in B$ and let $\varepsilon > 0$ be given.

We need to show that there is $\delta > 0$ such that, for $x \in B$,

$$\left|f|_B(x) - f|_B(a)\right| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta. \tag{3.1}$$

Since f is continuous at a, we know that there is a number $\delta > 0$ such that, for $x \in A$,

f is continuous at a because $a \in A$.

$$|f(x) - f(a)| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta.$$

In particular, since $B \subseteq A$, we know that, when $x \in B$ and $|x - a| < \delta$,

$$\left|f|_B(x) - f|_B(a)\right| = |f(x) - f(a)| < \varepsilon.$$

Thus (3.1) holds and hence $f|_B$ is continuous at a.

Since a is an arbitrary element of B, we conclude that $f|_B$ is continuous on B. ■

Worked problem 3.5

Show that the function $g\colon [0,1] \to \mathbb{R}$ given by

$$g(x) = \sqrt{x}$$

is continuous on $[0,1]$.

Solution

The function $\sqrt{}$ is a basic continuous function on $[0,\infty)$. Now g is just the restriction of this function to the interval $[0,1]$. It follows from Theorem 3.1 that g is continuous on $[0,1]$. ■

We now restate and prove the Composition Rule for continuous functions.

> ***Theorem 3.2 Composition Rule***
> Let $A, B \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ and $g\colon B \to \mathbb{R}$ be continuous on A and B respectively, with $f(A) \subseteq B$. Then the composed function $g \circ f\colon A \to \mathbb{R}$ is continuous on A.

Proof Consider a general point $a \in A$. For clarity, we write $b = f(a)$ and $y = f(x)$ for $x \in A$. Note that both y and b are in B.

Since f is continuous at a, for any $\varepsilon_1 > 0$ we can find $\delta_1 > 0$ such that

$$|f(x) - f(a)| < \varepsilon_1 \quad \text{whenever} \quad |x - a| < \delta_1. \tag{3.2}$$

Since g is continuous on B, for any $\varepsilon_2 > 0$ we can find $\delta_2 > 0$ such that

$$|g(y) - g(b)| < \varepsilon_2 \quad \text{whenever} \quad |y - b| < \delta_2. \tag{3.3}$$

Now substitute $f(a)$ for b and $f(x)$ for y. Then (3.3) becomes

$$|g(f(x)) - g(f(a))| < \varepsilon_2 \quad \text{whenever} \quad |f(x) - f(a)| < \delta_2,$$

which we can write equivalently as

$$|(g \circ f)(x) - (g \circ f)(a)| < \varepsilon_2 \quad \text{whenever} \quad |f(x) - f(a)| < \delta_2. \tag{3.4}$$

We can combine statements (3.2) and (3.4) by taking $\varepsilon_1 = \delta_2$. Thus, for $x \in A$, if $|x - a| < \delta_1$ then $|f(x) - f(a)| < \varepsilon_1 = \delta_2$, whence $|(g \circ f)(x) - (g \circ f)(a)| < \varepsilon_2$. Thus, for any $\varepsilon_2 > 0$ and for $x \in A$,

$$|(g \circ f)(x) - (g \circ f)(a)| < \varepsilon_2 \quad \text{whenever} \quad |x - a| < \delta_1.$$

Both ε_1 and δ_2 have dropped out of the final statement.

This means that $g \circ f$ is continuous at a. ■

Worked problem 3.6

Let $h\colon \mathbb{R} \to \mathbb{R}$ be given by

$$h(x) = \sqrt{1 + \sin x}.$$

Use the Composition Rule to prove that h is continuous on $\mathbb{R}$.

Solution

The function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = 1 + \sin x$ is the sum of two basic continuous functions and so is continuous. The function $g\colon [0, \infty) \to \mathbb{R}$ given by $g(x) = \sqrt{x}$ is a basic continuous function. We observe that the image set of $\mathbb{R}$ under $1 + \sin x$ is the interval $[0, 2]$, which is a subset of the domain $[0, \infty)$ of g. Moreover, for $x \in \mathbb{R}$,

$$(g \circ f)(x) = \sqrt{1 + \sin x} = h(x).$$

Thus, by the Composition Rule, $h = g \circ f$ is continuous on $\mathbb{R}$. ■

Note that the solution to Worked problem 3.6 is the same whether we use the ε–δ definition of continuity or the sequential definition. The Combination Rules, Restriction Rule and Composition Rule hold for both definitions. This is because the two definitions are equivalent. We now prove this, in Subsection 3.4.

3.4 Proof of equivalence

In this subsection, we prove the equivalence of our two definitions of continuity for functions $f\colon A \to \mathbb{R}$ where $A \subseteq \mathbb{R}$.

This subsection is not assessed.

For convenience, we repeat the two definitions.

Let $A \subseteq \mathbb{R}$ and $f\colon A \to \mathbb{R}$, and let $a \in A$.

Sequential continuity: f is continuous at a if, for each sequence (x_n) in A for which $x_n \to a$ as $n \to \infty$, the sequence $(f(x_n))$ converges to $f(a)$.

ε–δ continuity: f is continuous at a if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that, for all $x \in A$,

$$|f(x) - f(a)| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta.$$

What do we mean by equivalence? If the first definition implies the second *and* if the second definition implies the first, then we say that the definitions are **equivalent**. Thus, equivalence of these two definitions means that *if a function is continuous in the sequential sense, then it is also continuous in the ε–δ sense, and conversely.*

Informally, what is the difference between the two definitions?

In the ε–δ definition, we require that when we are given an ε-interval about $f(a)$, then we can find a δ-interval about a such that, whenever x is in this δ-interval, $f(x)$ is in the ε-interval. In other words, points that are close to a are mapped by f to points that are close to $f(a)$.

In the sequential definition, we require that, as x_n approaches a, $f(x_n)$ approaches $f(a)$ through a discrete set of values. This seems a lesser requirement, until we realize that it must hold for *all* sequences (x_n) that converge to a.

To link the two definitions, we associate each term x_n of any sequence with a particular δ-interval containing it. Thus the difference between the two requirements is not so great; and this is the essence of the proof. We can also predict that showing ε–δ continuity implies sequential continuity should be easier than the converse.

ε–δ continuity implies sequential continuity

We assume that $f\colon A \to \mathbb{R}$ is ε–δ continuous at the point $a \in A$.

Let (x_n) be an *arbitrary* sequence in A that converges to a. We must show that $f(x_n) \to f(a)$.

Let $\varepsilon > 0$ be given. Then, since f is ε–δ continuous at a, we can choose $\delta > 0$ such that, for $x \in A$,

$$|f(x) - f(a)| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta.$$

Since (x_n) converges to a, we can choose $N \in \mathbb{N}$ such that

$$|x_n - a| < \delta \quad \text{for all } n > N.$$

The integer N exists by the definition of convergence. See Subsection 1.3.

Thus, for $n > N$, $|x_n - a| < \delta$, and so $|f(x_n) - f(a)| < \varepsilon$.

So $f(x_n) \to f(a)$ as $n \to \infty$.

This proves that f is sequentially continuous at a.

Sequential continuity implies ε–δ continuity

We prove the converse using a *proof by contradiction.* We consider a function f that is sequentially continuous, and *assume that it is not ε–δ continuous.* With this latter assumption, we obtain a contradiction. Thus the assumption must be false and hence f *is* ε–δ continuous.

So, *assume that f is sequentially continuous at $a \in A$, but that f is not ε–δ continuous at a.* Since f is *not* ε–δ continuous at a, there must be a particular $\varepsilon > 0$ such that in each δ-interval $(a-\delta, a+\delta)$ around a there is at least one point x for which $|f(x) - f(a)| \geq \varepsilon$.

The trick is not to start with an arbitrary ε, but to choose one that does the job.

Take $\delta = \frac{1}{n}$. Then there is at least one point in the interval $(a - \frac{1}{n}, a + \frac{1}{n})$ for which $|f(x) - f(a)| \geq \varepsilon$. Call this point x_n. Now, for each value of $n \in \mathbb{N}$, there is such a point. We know that $a - \frac{1}{n} < x_n < a + \frac{1}{n}$, so that $-\frac{1}{n} < x_n - a < \frac{1}{n}$, i.e. $|x_n - a| < \frac{1}{n}$. These x_n form a sequence (x_n) such that $x_n \to a$. This is because $|x_n - a| < \frac{1}{n}$ for all n and $(\frac{1}{n})$ is a null sequence; hence $(x_n - a)$ is a null sequence, by the Squeeze Rule. But $f(x_n) \not\to f(a)$, since $|f(x_n) - f(a)| \geq \varepsilon$ for all n. This contradicts the fact that f is sequentially continuous at a.

Since we have obtained a contradiction, our initial assumption that f is not ε–δ continuous must be wrong. Thus, sequential continuity implies ε–δ continuity.

It follows that the two definitions of continuity are equivalent.

4 Continuity on the plane

After working through this section, you should be able to:

- calculate the *distance* between any two points in the plane $\mathbb{R}^2$;
- explain the definition of *continuity* for functions from $\mathbb{R}^2$ to $\mathbb{R}$;
- determine whether a given function from $\mathbb{R}^2$ to $\mathbb{R}$ is continuous;
- understand the *Euclidean distance function* on $\mathbb{R}^2$, and appreciate its properties.

One of the main ways in which mathematics develops is by taking ideas from one situation and extending or generalizing them into new ones. This process involves isolating some key aspects and ignoring others. In this section we move from functions on the real line to functions defined on the plane: the domain of these functions is the plane, $\mathbb{R}^2$, and their codomain is $\mathbb{R}$. We shall extend the ideas of distance and continuity which we have developed in Sections 1–3 to make them appropriate for functions from $\mathbb{R}^2$ to $\mathbb{R}$.

In Section 5 we continue this process by extending these ideas to spaces of higher dimensions. In *Unit A2* we consider more general ideas of distance.

4.1 Functions from $\mathbb{R}^2$ to $\mathbb{R}$

In $\mathbb{R}^2$, we work with the familiar coordinate geometry of two dimensions. However, instead of the common notation (x, y) for points in the plane, we use the notation (x_1, x_2) since this subscript notation extends more easily to spaces of higher dimensions, which we consider in Section 5. We also use the notation $\mathbf{x}$ for the point (x_1, x_2), and $\mathbf{0}$ for the point $(0, 0)$. We refer to the point $\mathbf{0}$ as the **origin** of $\mathbb{R}^2$. For any two points $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ in $\mathbb{R}^2$, we define

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2) \quad \text{and} \quad \mathbf{a} - \mathbf{b} = (a_1 - b_1, a_2 - b_2).$$

In specifying the rule of a function f with domain $\mathbb{R}^2$ we use the notation $f(\mathbf{x})$ or $f(x_1, x_2)$. Two examples of functions from $\mathbb{R}^2$ to $\mathbb{R}$ are:

$$f(x_1, x_2) = 20 - (x_1^2 + x_2^2) \quad \text{and} \quad g(x_1, x_2) = \frac{x_1^2 x_2}{x_1^2 + x_2^2 + 1}.$$

Although, strictly speaking, $f(\mathbf{x}) = f((x_1, x_2))$, we omit one pair of brackets and write $f(x_1, x_2)$.

Just as we represent functions from $\mathbb{R}$ to $\mathbb{R}$ by curves in the plane, so we can represent functions from $\mathbb{R}^2$ to $\mathbb{R}$ by *surfaces* in three-dimensional space. We use a three-dimensional coordinate system: above each point (x_1, x_2) in the domain $\mathbb{R}^2$, which we draw as a horizontal plane, we mark a point whose third coordinate x_3 is the function value, $f(x_1, x_2)$. Thus, above the point $(3, 2)$ we put $f(3, 2) = 20 - 13 = 7$. As (x_1, x_2) ranges over different values in the domain $\mathbb{R}^2$, the values of $x_3 = f(x_1, x_2)$ define a surface in space. The graphs of the functions f and g are shown as surfaces in three-dimensional space in Figure 4.1.

Problem 4.1

(a) Calculate $f(1, -2)$ and $g(-1, 3)$, and locate these points approximately on their graphs.

(b) Describe those points on the graph of g for which $x_1 = 0$.

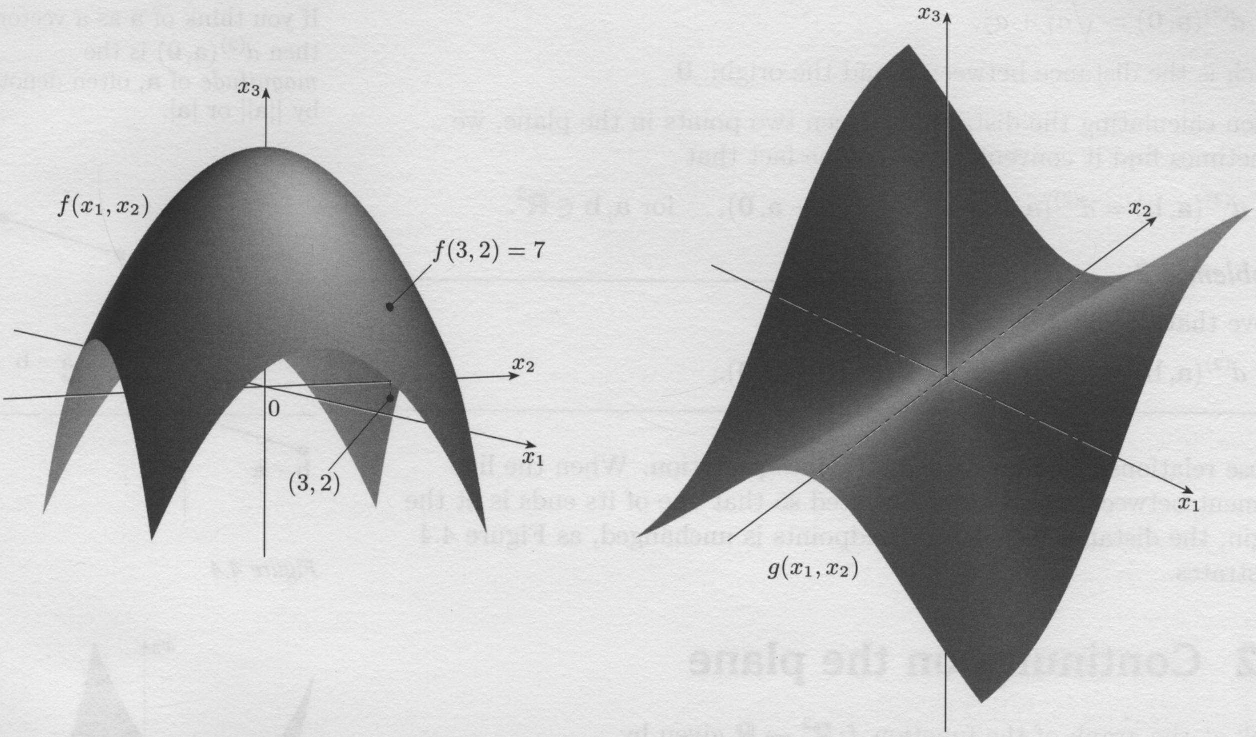


Figure 4.1

As a first step in considering the continuity of functions on $\mathbb{R}^2$, we need to extend the definition of distance that we used for functions on $\mathbb{R}$. We saw in Section 1 that, on the real line $\mathbb{R}$, the distance between points a and b is given by $|b - a|$. In $\mathbb{R}^2$ we want the distance between two points to be defined in the usual way: for example, as Figure 4.2 illustrates, the distance between the points $(3, -1)$ and $(7, 2)$ is

$$\sqrt{(7-3)^2 + (2-(-1))^2} = \sqrt{4^2 + 3^2} = 5.$$

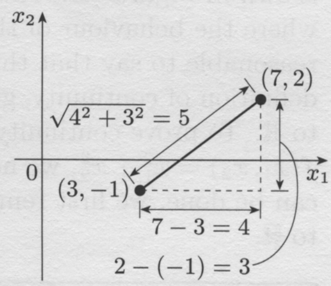


Figure 4.2

This distance, which is calculated by Pythagoras's Theorem, is called the *Euclidean distance* in $\mathbb{R}^2$.

Definition

The **Euclidean distance** between points $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ in $\mathbb{R}^2$ is given by the formula

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

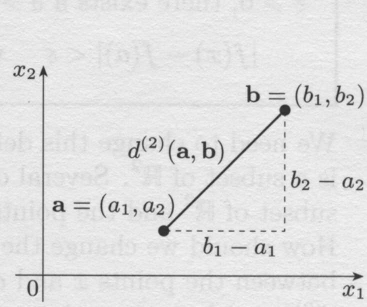


Figure 4.3

Remark

Notice the use of the superscript $^{(2)}$ in $d^{(2)}(\mathbf{a}, \mathbf{b})$. It denotes that this is a distance between points in the plane — that is, in *two* dimensions. This notation is extended to higher dimensions in Section 5.

For example, $d^{(2)}((3, -1), (7, 2)) = \sqrt{(7-3)^2 + (2-(-1))^2} = 5$.

Putting $\mathbf{b} = \mathbf{0} = (0, 0)$ in the definition of $d^{(2)}(\mathbf{a}, \mathbf{b})$ gives

$$d^{(2)}(\mathbf{a}, \mathbf{0}) = \sqrt{a_1^2 + a_2^2},$$

which is the distance between $\mathbf{a}$ and the origin, $\mathbf{0}$.

If you think of $\mathbf{a}$ as a vector, then $d^{(2)}(\mathbf{a}, \mathbf{0})$ is the *magnitude* of $\mathbf{a}$, often denoted by $||\mathbf{a}||$ or $|\mathbf{a}|$.

When calculating the distance between two points in the plane, we sometimes find it convenient to use the fact that

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0}), \quad \text{for } \mathbf{a}, \mathbf{b} \in \mathbb{R}^2.$$

Problem 4.2

Prove that, for all $\mathbf{a}$, $\mathbf{b}$ in $\mathbb{R}^2$,

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0}).$$

These relationships have a geometric interpretation. When the line segment between $\mathbf{a}$ and $\mathbf{b}$ is translated so that one of its ends is at the origin, the distance between its endpoints is unchanged, as Figure 4.4 illustrates.

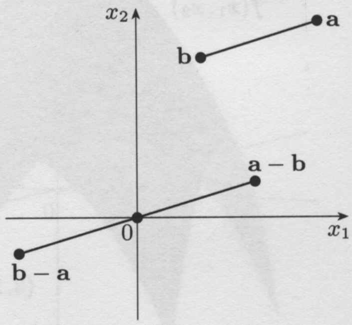


Figure 4.4

4.2 Continuity on the plane

Look at the graph of the function $f: \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x_1, x_2) = x_1^2 + x_2^2,$$

shown in Figure 4.5. There are no obvious jumps or gaps, and no regions where the behaviour of the function is strange. It certainly seems reasonable to say that this function *is continuous* on $\mathbb{R}^2$. However, the ε–δ definition of continuity, given in Section 3, applies only to functions from $\mathbb{R}$ to $\mathbb{R}$. To prove continuity for functions from $\mathbb{R}^2$ to $\mathbb{R}$, such as $f(x_1, x_2) = x_1^2 + x_2^2$, we need to extend that definition. To show how this can be done, we first remind you of the ε–δ definition of continuity from $\mathbb{R}$ to $\mathbb{R}$.

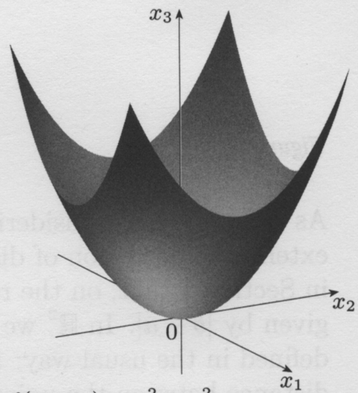


Figure 4.5

> Let $A \subseteq \mathbb{R}$. A function $f: A \to \mathbb{R}$ is **continuous** at $a \in A$ if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that, for all $x \in A$,
>
> $$|f(x) - f(a)| < \varepsilon \quad \text{whenever} \quad \boxed{|x - a| < \delta}.$$

We need to change this definition to refer to the fact that the domain of f is a subset of $\mathbb{R}^2$. Several of the changes are straightforward: A becomes a subset of $\mathbb{R}^2$ and the points a and x in $\mathbb{R}$ become points $\mathbf{a}$ and $\mathbf{x}$ in $\mathbb{R}^2$. How should we change the boxed item? Since $|x - a|$ is the distance between the points x and a in $\mathbb{R}$, we need to change this to the distance $d^{(2)}(\mathbf{x}, \mathbf{a})$ between points $\mathbf{x}$ and $\mathbf{a}$ in the plane. These changes lead to the following definition of continuity.

> ***Definition***
>
> Let $A \subseteq \mathbb{R}^2$. A function $f: A \to \mathbb{R}$ is **continuous** at $\mathbf{a} \in A$ if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that, for all $\mathbf{x} \in A$,
>
> $$|f(\mathbf{x}) - f(\mathbf{a})| < \varepsilon \quad \text{whenever} \quad d^{(2)}(\mathbf{x}, \mathbf{a}) < \delta.$$

Remarks

(i) Informally, this definition says that, given a number $\varepsilon > 0$, we must find a number $\delta > 0$ such that, if $\mathbf{x}$ is within a distance δ of $\mathbf{a}$, then $f(\mathbf{x})$ lies within a distance ε of $f(\mathbf{a})$. This is illustrated in Figure 4.6.
In functions from $\mathbb{R}$ to $\mathbb{R}$, the condition $|x - a| < \delta$ (x must be within a distance δ of a) gives rise to an open δ-interval on the real line.
Here the corresponding condition $d^{(2)}(\mathbf{x}, \mathbf{a}) < \delta$ ($\mathbf{x}$ must be within a distance δ of $\mathbf{a}$) gives an open disc of radius δ, centred at $\mathbf{a}$. In Figure 4.6, the open disc is the shaded circular region on the (x_1, x_2)-plane.

An *open disc* in $\mathbb{R}^2$ corresponds to an open interval in $\mathbb{R}$. Just as an open interval excludes its endpoints, so an open disc excludes its boundary points.

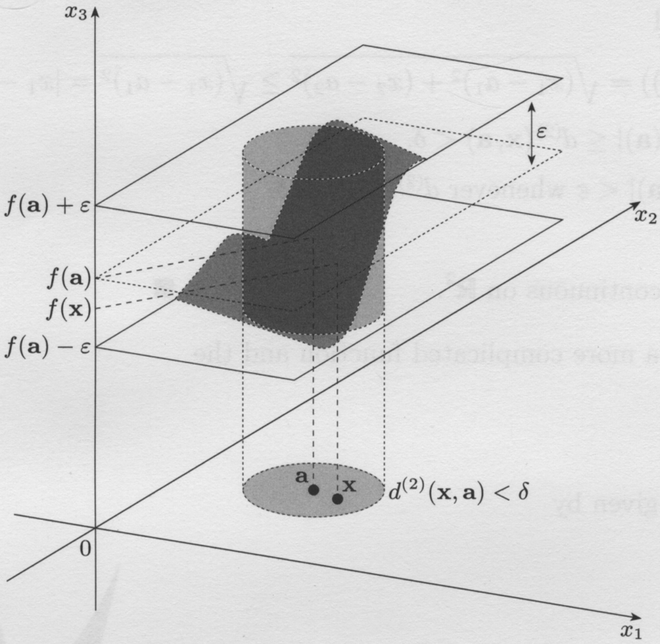


Figure 4.6

(ii) As before, the value of δ may depend both on ε and on the point $\mathbf{a}$ (that is, on a_1 and a_2).

Definition

Let $S \subseteq A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ be a function. We say that f is **continuous** on S when f is continuous at each point in S.

Our first examples of continuous functions from the plane to the real line are very simple but will prove useful later.

Definition

The two functions $p_1\colon \mathbb{R}^2 \to \mathbb{R}$, $p_2\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$p_1(x_1, x_2) = x_1, \quad p_2(x_1, x_2) = x_2,$$

are known as **projection functions**.

Remark

These functions *project* each point $\mathbf{a} = (a_1, a_2)$ in $\mathbb{R}^2$ onto the coordinate axes, as Figure 4.7 illustrates.

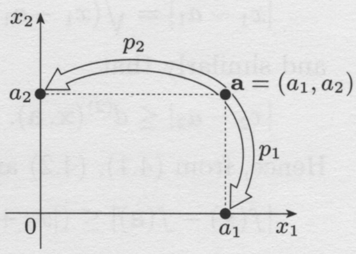


Figure 4.7

Worked problem 4.1

Prove that the projection functions p_1 and p_2 are continuous on $\mathbb{R}^2$.

Solution

We give the proof for p_1. The proof for p_2 is similar.

Let $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$ be a general point in the plane, and let $\varepsilon > 0$ be given.

We have to find $\delta > 0$ such that, for $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$,

$$|p_1(\mathbf{x}) - p_1(\mathbf{a})| < \varepsilon \quad \text{whenever} \quad d^{(2)}(\mathbf{x}, \mathbf{a}) < \delta.$$

Now $|p_1(\mathbf{x}) - p_1(\mathbf{a})| = |x_1 - a_1|$ and

$$d^{(2)}(\mathbf{x}, \mathbf{a}) = d^{(2)}((x_1, x_2), (a_1, a_2)) = \sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2} \geq \sqrt{(x_1 - a_1)^2} = |x_1 - a_1|.$$

So, when $d^{(2)}(\mathbf{x}, \mathbf{a}) < \delta$, $|p_1(\mathbf{x}) - p_1(\mathbf{a})| \leq d^{(2)}(\mathbf{x}, \mathbf{a}) < \delta$.

Thus, by taking $\delta = \varepsilon$, $|p_1(\mathbf{x}) - p_1(\mathbf{a})| < \varepsilon$ whenever $d^{(2)}(\mathbf{x}, \mathbf{a}) < \delta$.

Hence p_1 is continuous at $\mathbf{a}$.

Since $\mathbf{a}$ is an arbitrary point, p_1 is continuous on $\mathbb{R}^2$. ■

The next worked problem involves a more complicated function and the solution requires some extra effort.

Worked problem 4.2

Prove that the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x_1, x_2) = x_1^2 + x_2^2$$

is continuous on $\mathbb{R}^2$.

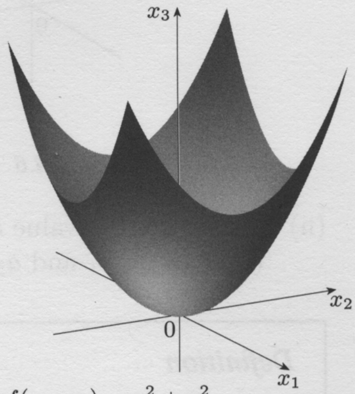


Figure 4.8

Solution

Let $\mathbf{a} \in \mathbb{R}^2$ and let $\varepsilon > 0$ be given.

We have to find $\delta > 0$ such that, for $\mathbf{x} \in \mathbb{R}^2$,

$$|f(\mathbf{x}) - f(\mathbf{a})| < \varepsilon \quad \text{whenever} \quad d^{(2)}(\mathbf{x}, \mathbf{a}) < \delta.$$

First, we seek an upper bound for $|f(\mathbf{x}) - f(\mathbf{a})|$ in terms of $d^{(2)}(\mathbf{x}, \mathbf{a})$: we use the Triangle Inequality.

$$\begin{aligned} |f(\mathbf{x}) - f(\mathbf{a})| &= |(x_1^2 + x_2^2) - (a_1^2 + a_2^2)| \\ &= |(x_1^2 - a_1^2) + (x_2^2 - a_2^2)| \\ &\leq |x_1^2 - a_1^2| + |x_2^2 - a_2^2| \qquad \text{(by the Triangle Inequality)} \\ &= |x_1 - a_1|\,|x_1 + a_1| + |x_2 - a_2|\,|x_2 + a_2|. \qquad (4.1) \end{aligned}$$

We now use the fact that

$$|x_1 - a_1| = \sqrt{(x_1 - a_1)^2} \leq \sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2} = d^{(2)}(\mathbf{x}, \mathbf{a}), \qquad (4.2)$$

and similarly that

$$|x_2 - a_2| \leq d^{(2)}(\mathbf{x}, \mathbf{a}). \qquad (4.3)$$

Hence, from (4.1), (4.2) and (4.3),

$$|f(\mathbf{x}) - f(\mathbf{a})| \leq (|x_1 + a_1| + |x_2 + a_2|) d^{(2)}(\mathbf{x}, \mathbf{a}). \qquad (4.4)$$

It still remains to find upper bounds for $|x_1 + a_1|$ and $|x_2 + a_2|$. To do this for $|x_1 + a_1|$, we rewrite it as $|(x_1 - a_1) + 2a_1|$ and use the triangle

inequality and (4.2) to obtain

$$|x_1 + a_1| = |(x_1 - a_1) + 2a_1| \leq |x_1 - a_1| + 2|a_1|$$
$$\leq d^{(2)}(\mathbf{x}, \mathbf{a}) + 2|a_1|. \tag{4.5}$$

Similarly, for $|x_2 + a_2|$,

$$|x_2 + a_2| \leq d^{(2)}(\mathbf{x}, \mathbf{a}) + 2|a_2|. \tag{4.6}$$

Hence, from (4.4), (4.5) and (4.6),

$$|f(\mathbf{x}) - f(\mathbf{a})| \leq \left(2d^{(2)}(\mathbf{x}, \mathbf{a}) + 2|a_1| + 2|a_2|\right) d^{(2)}(\mathbf{x}, \mathbf{a}).$$

Thus, whenever $d^{(2)}(\mathbf{x}, \mathbf{a}) < \delta$, we have

$$|f(\mathbf{x}) - f(\mathbf{a})| < 2(\delta + |a_1| + |a_2|)\delta.$$

We now choose $\delta > 0$ such that

$$2(\delta + |a_1| + |a_2|)\delta \leq \varepsilon.$$

We can assume that $\delta \leq 1$. It is then sufficient to find δ for which

$$2(1 + |a_1| + |a_2|)\delta \leq \varepsilon.$$

We thus set $\delta = \min\left\{1, \dfrac{\varepsilon}{2(1 + |a_1| + |a_2|)}\right\}$.

Hence, $|f(\mathbf{x}) - f(\mathbf{a})| < \varepsilon$ whenever $d^{(2)}(\mathbf{x}, \mathbf{a}) < \delta$.

Therefore f is continuous at $\mathbf{a}$.

Since $\mathbf{a}$ is an arbitrary point, f is continuous on $\mathbb{R}^2$. ■

As Worked problem 4.2 illustrates, it would be tedious if we had to work directly from the definition of continuity whenever we had to prove that a function from $\mathbb{R}^2$ to $\mathbb{R}$ is continuous. Fortunately, as with functions from $\mathbb{R} \to \mathbb{R}$, there are Combination Rules that allow us to combine continuous functions to form new continuous functions. We state them without proof.

Combination Rules for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$

Let $A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ and $g\colon A \to \mathbb{R}$ be continuous on A.

Then the following functions are continuous on A:

Sum $f + g\colon A \to \mathbb{R}$, defined by $(f + g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$;

Multiple $\lambda f\colon A \to \mathbb{R}$ where $\lambda \in \mathbb{R}$, defined by $(\lambda f)(\mathbf{x}) = \lambda \times f(\mathbf{x})$;

Product $fg\colon A \to \mathbb{R}$, defined by $(fg)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$.

Worked problem 4.3

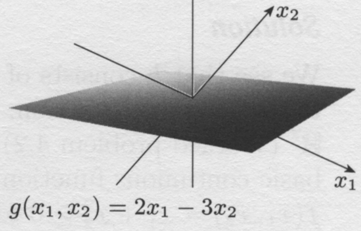
Let $\mathbf{u} = (u_1, u_2)$ be a fixed point in $\mathbb{R}^2$. Use the Combination Rules to prove that the function $g\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$g(\mathbf{x}) = u_1 x_1 + u_2 x_2$$

is continuous on $\mathbb{R}^2$.

Remark

The graph of g is a plane through the origin, as Figure 4.9 illustrates for $\mathbf{u} = (2, -3)$. It is intuitively clear that g is continuous everywhere, but this still needs proof.



$g(x_1, x_2) = 2x_1 - 3x_2$

Figure 4.9

Solution

The key to proving that this function is continuous is to recognize that g can be written in terms of the projection functions p_1 and p_2. In fact,

$$g(\mathbf{x}) = u_1 p_1(\mathbf{x}) + u_2 p_2(\mathbf{x}).$$

Since both p_1 and p_2 are continuous on $\mathbb{R}^2$ and since $\mathbf{u} = (u_1, u_2)$ is fixed, the Multiple Rule guarantees that $u_1 p_1$ and $u_2 p_2$ are both continuous functions on $\mathbb{R}^2$. We can now apply the Sum Rule to $u_1 p_1$ and $u_2 p_2$ to deduce that $g = u_1 p_1 + u_2 p_2$ is continuous on $\mathbb{R}^2$. ■

The result of Worked problem 4.3 applies for any $\mathbf{u}$; so it tells us, for example, that the function $f(\mathbf{x}) = 2x_1 - 3x_2$, which is $g(\mathbf{x})$ with $\mathbf{u} = (2, -3)$, is continuous on $\mathbb{R}^2$.

There is also a Restriction Rule for functions from $\mathbb{R}^2$ to $\mathbb{R}$ which allows us to restrict the domain of a continuous function to a smaller set and still have a continuous function. The proof is similar to that for functions from $\mathbb{R}$ to $\mathbb{R}$ (see Theorem 3.1), so we omit it.

> ***Theorem 4.1 Restriction Rule***
>
> Let $A \subseteq \mathbb{R}^2$, let $f\colon A \to \mathbb{R}$ be continuous on A and let $B \subseteq A$. Then the restricted function $f|_B : B \to \mathbb{R}$ is continuous on B.

Another useful rule is the one that allows us to compose continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ with continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Again we state it without proof.

The proof is similar to that for functions from $\mathbb{R}$ to $\mathbb{R}$ (see Theorem 3.2).

> ***Theorem 4.2 Composition Rule***
>
> Let $A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ be continuous on A. Let $f(A) \subseteq B \subseteq \mathbb{R}$ and let $g : B \to \mathbb{R}$ be continuous on B. Then the composed function $g \circ f\colon A \to \mathbb{R}$ is continuous on A.

Remark

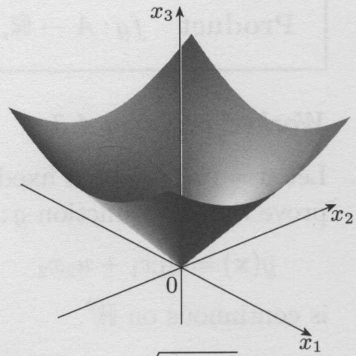
Notice that we are *not* composing two functions from $\mathbb{R}^2$ to $\mathbb{R}$. Since the codomain of f is $\mathbb{R}$, and not $\mathbb{R}^2$, the domain of g must be a subset of $\mathbb{R}$.

Worked problem 4.4

Use the rules to prove that the function $h\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$h(\mathbf{x}) = \sqrt{x_1^2 + x_2^2}$$

is continuous on $\mathbb{R}^2$.



$h(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$

Figure 4.10

Solution

We see that h consists of composing the function $f(x_1, x_2) = x_1^2 + x_2^2$ with the square-root function. We have already shown that f is continuous on $\mathbb{R}^2$ (Worked problem 4.2) and we know that the square-root function is a basic continuous function from $[0, \infty)$ to $\mathbb{R}$. Furthermore, since $f(x_1, x_2) = x_1^2 + x_2^2 \geq 0$ for all real numbers, the image set of f is included in the domain $[0, \infty)$ of the square-root function. Hence the conditions of the Composition Rule are satisfied, so we conclude that h is continuous on $\mathbb{R}^2$. ■

Problem 4.3

Use the projection functions and the rules to prove that the function $f: \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\mathbf{x}) = \sin(x_1 + x_2)$$

is continuous on $\mathbb{R}^2$.

Problem 4.4

Use the projection functions and the rules to prove that the function $f: \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\mathbf{x}) = e^{-x_1 x_2}$$

is continuous on $\mathbb{R}^2$.

What about discontinuous functions on $\mathbb{R}^2$? Recall the function $f_2: \mathbb{R} \to \mathbb{R}$ given by

$$f_2(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

We saw that f_2 is not continuous at 0 because f_2 approaches different limiting values (1 and 0) as x approaches 0 from above or below. There is an analogous situation for functions whose domains are subsets of $\mathbb{R}^2$. If a function approaches different limits when a point is approached from different directions, then the function is discontinuous at the point.

See Worked problems 2.2 and 3.2.

Formally, the definition of a discontinuous function on $\mathbb{R}^2$ is the same as that on $\mathbb{R}$ but with $\mathbb{R}$ replaced by $\mathbb{R}^2$.

Definition

Let $S \subseteq A \subseteq \mathbb{R}^2$ and let $f: A \to \mathbb{R}$ be a function.

Then f is **discontinuous** at $\mathbf{a} \in A$ if f is not continuous at $\mathbf{a}$.

Furthermore, f is **discontinuous** on S if f is discontinuous at at least one point in S.

To show that a function from $\mathbb{R}^2$ to $\mathbb{R}$ is discontinuous by ε–δ methods, we need to find *just one value* of ε such that, for *all* $\delta > 0$, we can find an $\mathbf{x} \in \mathbb{R}^2$ for which

$$d^{(2)}(\mathbf{x}, \mathbf{a}) < \delta \quad \text{and} \quad |f(\mathbf{x}) - f(\mathbf{a})| \geq \varepsilon.$$

Worked problem 4.5

Show that the function $f: \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\mathbf{x}) = \begin{cases} \dfrac{x_1^2 - x_2^2}{x_1^2 + x_2^2} & \text{if } \mathbf{x} \neq \mathbf{0}, \\ 0 & \text{if } \mathbf{x} = \mathbf{0}, \end{cases}$$

is discontinuous at $\mathbf{0}$.

Solution

The graph of the function is shown in Figure 4.11. Look along the positive x_1-axis, where $x_2 = 0$. There, $f(\mathbf{x}) = f(x_1, 0) = 1$, except at the origin $\mathbf{0}$, where $f(\mathbf{0}) = 0$.

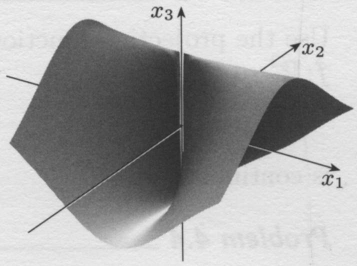


Figure 4.11

In order to show discontinuity at $\mathbf{0}$, we need to find just one value of $\varepsilon > 0$ such that, for all $\delta > 0$, we can find an $\mathbf{x} \in \mathbb{R}^2$ for which

$$d^{(2)}(\mathbf{x}, \mathbf{0}) < \delta \quad \text{and} \quad |f(\mathbf{x}) - f(\mathbf{0})| \geq \varepsilon.$$

Using the behaviour of f along the x_1-axis as our guide, we choose $\varepsilon = \frac{1}{2}$. Then, for all $\delta > 0$, the point $\mathbf{x} = (\frac{1}{2}\delta, 0)$ has $d^{(2)}(\mathbf{x}, \mathbf{0}) = \frac{1}{2}\delta < \delta$. But, at this point, $f(\mathbf{x}) = 1$ and so

Any ε with $0 < \varepsilon \leq 1$ would do, and any point $\mathbf{x} = (x_1, 0)$ with $0 < x_1 < \delta$ would do.

$$|f(\mathbf{x}) - f(\mathbf{0})| = |1 - 0| = 1 \geq \varepsilon.$$

Hence f is discontinuous at $\mathbf{0}$. ■

Remark

Along the line $x_1 = x_2$, $f(\mathbf{x})$ is always 0. Although this might suggest that f is continuous at $\mathbf{0}$, this is not the case. To be continuous at $\mathbf{0}$, it is necessary that $f(\mathbf{x})$ must approach the same value *whatever direction is taken.*

Problem 4.5

Show that the function $g : \mathbb{R}^2 \to \mathbb{R}$ given by

$$g(x_1, x_2) = \begin{cases} 0 & \text{if } x_2 \geq x_1, \\ 1 & \text{if } x_2 < x_1, \end{cases}$$

is discontinuous at $(2, 2)$.

4.3 Properties of distance

Because we wish to generalize the idea of distance to higher dimensions, and to other contexts, we need to highlight aspects that we have not considered before.

So far, we have met two ways of defining distance: the distance between points a and b on the real line is $|b - a|$, and the distance between points $\mathbf{a}$ and $\mathbf{b}$ in the plane is

Another definition of distance with which you may be familiar is the great circle distance between points on the surface of a sphere. You will meet other kinds of distance in *Unit A2.*

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

These distances can be thought of as functions of two variables. On the real line, we associate with two numbers a and b a single non-negative real number $|b - a|$. Similarly, on the plane, we associate with two points (a_1, a_2) and (b_1, b_2) the single non-negative real number

$$\sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Both functions are referred to as *Euclidean distance*, on $\mathbb{R}$ and on $\mathbb{R}^2$ respectively.

Each of these distance functions has three significant properties, the first two of which may seem almost trivial. We draw attention to these three properties because they are the key properties that any definition of distance must satisfy. We describe these properties for the Euclidean distance function $d^{(2)}$ on $\mathbb{R}^2$.

The first property is that the distance between two *distinct* points is always positive, and the distance between two points that coincide is 0. This property is labelled (M1) and is stated formally as follows.

The reason for the M in the labels (M1), (M2) and (M3) will become clear in the next unit.

(M1) For all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, $d^{(2)}(\mathbf{a}, \mathbf{b}) \geq 0$, with $d^{(2)}(\mathbf{a}, \mathbf{b}) = 0$ if and only if $\mathbf{a} = \mathbf{b}$.

The second property is that it does not matter in which direction you measure the distance between two points: the distance from $\mathbf{a}$ to $\mathbf{b}$ is the same as the distance from $\mathbf{b}$ to $\mathbf{a}$. We say that $d^{(2)}$ is *symmetric*.

(M2) For all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, $d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{b}, \mathbf{a})$.

Worked problem 4.6

Prove that the distance property (M2) holds.

Solution

This follows from the fact that, for any $p, q \in \mathbb{R}$, $(p-q)^2 = (q-p)^2$.

Let $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$. Then

$$\begin{aligned} d^{(2)}(\mathbf{a}, \mathbf{b}) &= \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2} \\ &= \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} = d^{(2)}(\mathbf{b}, \mathbf{a}). \end{aligned}$$ ■

The third important property is the **Triangle Inequality**. This inequality expresses the fact that, in a triangle, the length of any side cannot exceed the sum of the lengths of the other two sides. It can be expressed symbolically as follows, and is illustrated in Figure 4.12.

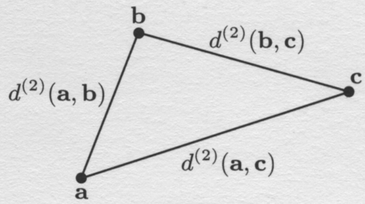


Figure 4.12

(M3) For all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^2$, $d^{(2)}(\mathbf{a}, \mathbf{c}) \leq d^{(2)}(\mathbf{a}, \mathbf{b}) + d^{(2)}(\mathbf{b}, \mathbf{c})$.

Problem 4.6

Verify the Triangle Inequality for the points

$$\mathbf{a} = (3, 4), \quad \mathbf{b} = (6, 0), \quad \mathbf{c} = (5, -3).$$

There are many proofs of the Triangle Inequality. We prove a general version for n dimensions in Subsection 5.4. The case when $n = 2$ provides a proof of the Triangle Inequality in two dimensions.

The three properties (M1)–(M3) also apply to the Euclidean distance function $|b-a|$ on $\mathbb{R}$, and were demonstrated in Subsection 1.2. They can be summarized as follows.

(M1) For all $a, b \in \mathbb{R}$, $|b-a| \geq 0$, with $|b-a| = 0$ if and only if $b = a$.

(M2) For all $a, b \in \mathbb{R}$, $|b-a| = |a-b|$.

(M3) For all $a, b, c \in \mathbb{R}$, $|c-a| \leq |b-a| + |c-b|$.

Remark

The version of the Triangle Inequality for $\mathbb{R}$ given in Subsection 1.2 is $|a+b| \leq |a| + |b|$. This is equivalent to (M3), as we saw in Problem 1.7.

The properties (M1)–(M3) hold not only for the Euclidean distance functions defined on $\mathbb{R}$ and $\mathbb{R}^2$, but also for Euclidean distance in higher dimensions, as we show in Section 5. In fact, these three properties define the notion of 'distance' completely. In *Unit A2*, we shall specify that *any* set endowed with a function satisfying properties (M1)–(M3) is a set in which distance is defined. In that way we shall obtain a deeper understanding of distance.

5 Continuity in higher dimensions

After working through this section, you should be able to:

- understand the *Euclidean distance function* on $\mathbb{R}^n$, and appreciate its properties;
- state and use the *Triangle* and *Reverse Triangle Inequalities* for the Euclidean distance function on $\mathbb{R}^n$;
- explain the definition of *continuity* for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$;
- determine whether a given function from $\mathbb{R}^n$ to $\mathbb{R}^m$ is continuous.

In this section we continue our process of abstracting the ideas of distance and of the continuity of functions. Having dealt with distance and continuity in one and two dimensions, we now look at them in higher dimensions.

Here we examine distance and the continuity of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. The methods we adopt are similar to those already used, with some extra complications because of the generality involved. You will find that results such as the definition of continuity, the Combination Rules and the Triangle Inequality have close analogies with the results in lower dimensions. Since we cannot visualize distances and functions in such spaces, we have to rely on algebraic methods to carry through these extensions.

Replacing $\mathbb{R}^n$ by $\mathbb{R}^3$ can sometimes help you to understand general results.

5.1 Functions and distance

First, we need to generalize the ideas of function and distance to higher dimensions. We consider a space $\mathbb{R}^n$ of general dimension n. Points in $\mathbb{R}^n$ have n coordinates, and are denoted by ordered *n-tuples* $(x_1, x_2, \ldots, x_n)$. As in two dimensions, we use the notation $\mathbf{x}$ for the point $(x_1, x_2, \ldots, x_n)$ and $\mathbf{0}$ for the point $(0, 0, \ldots, 0)$, the **origin** of $\mathbb{R}^n$. Also, if $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$, then

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n),$$
$$\mathbf{a} - \mathbf{b} = (a_1 - b_1, a_2 - b_2, \ldots, a_n - b_n).$$

Some examples of functions in higher dimensions are:

- $f: \mathbb{R}^4 \to \mathbb{R}$ given by $f(x_1, x_2, x_3, x_4) = x_1 + x_2 + x_3 + x_4$, so that, for instance, $f(1, -4, 0, 2) = -1$;
- $g: \mathbb{R}^3 \to \mathbb{R}^2$ given by $g(x_1, x_2, x_3) = (x_3, x_2)$, so that, for instance, $g(3, 7, 1) = (1, 7)$;
- $h: \mathbb{R}^5 \to \mathbb{R}^3$ given by $h(x_1, x_2, x_3, x_4, x_5) = (2x_1, x_2^2 + 3x_4, x_3 + x_5)$, so that, for instance, $h(1, -2, 3, -3, 4) = (2, -5, 7)$;
- $k: \mathbb{R}^2 \to \mathbb{R}^3$ given by $k(x_1, x_2) = (\cos 2\pi x_1, \sin \pi x_2, x_1 \sin 2\pi x_2)$, so that, for instance, $k(1, 1.5) = (1, -1, 0)$.

Notice that the domains and the codomains of these functions can have different dimensions. In general, we consider functions from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Problem 5.1

Evaluate the following functions at the given points:

(a) $g : \mathbb{R}^3 \to \mathbb{R}^2, g(x_1, x_2, x_3) = (x_3, x_2)$ at $(2, -1, 0)$;

(b) $k : \mathbb{R}^2 \to \mathbb{R}^3, k(x_1, x_2) = (\cos 2\pi x_1, \sin \pi x_2, x_1 \sin \pi x_2)$ at $(0, 0.5)$.

In Section 4 we generalized the definition of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ to functions from $\mathbb{R}^2$ to $\mathbb{R}$. An essential observation was that the modulus function acts as the distance function on $\mathbb{R}$, which led us to consider the distance function $d^{(2)}$ on the plane. In order to define continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, we need to generalize the Euclidean distance formula to $\mathbb{R}^n$.

Definition

The **Euclidean distance** between points $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in $\mathbb{R}^n$ is given by the formula

$$d^{(n)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2}.$$

Remarks

(i) This distance formula is consistent with those for $\mathbb{R}$ and $\mathbb{R}^2$.
When $n = 1$, we obtain $d^{(1)}(a, b) = \sqrt{(b-a)^2} = |b - a|$.
When $n = 2$, we obtain the distance between points in the plane.

(ii) As with $d^{(2)}$, we have $d^{(n)}(\mathbf{a}, \mathbf{b}) = d^{(n)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(n)}(\mathbf{b} - \mathbf{a}, \mathbf{0})$.

For example, in $\mathbb{R}^4$,

$$\begin{aligned} d^{(4)}((1,2,3,4),(2,-1,0,3)) &= \sqrt{(2-1)^2 + (-1-2)^2 + (0-3)^2 + (3-4)^2} \\ &= \sqrt{1+9+9+1} = \sqrt{20} = 4.472\ldots. \end{aligned}$$

Problem 5.2

In $\mathbb{R}^5$, calculate the distance $d^{(5)}((4, 2, 3, 0, -1), (2, -3, -1, 0, 2))$.

5.2 Properties of the distance function

In Subsection 4.3 we gave three properties possessed by the Euclidean distance functions on $\mathbb{R}$ and on $\mathbb{R}^2$. Here we verify that the Euclidean distance function $d^{(n)}(\mathbf{a}, \mathbf{b})$ also satisfies these properties.

The first property is that distances are never negative and that the distance between two points is zero if and only if the two points are the same.

(M1) For all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, $d^{(n)}(\mathbf{a}, \mathbf{b}) \geq 0$, with $d^{(n)}(\mathbf{a}, \mathbf{b}) = 0$ if and only if $\mathbf{a} = \mathbf{b}$.

This is referred to as the *non-negativity property*.

The property that distances are never negative is built into $d^{(n)}$ through the square root. To show that the second part of (M1) holds, suppose first that $\mathbf{a} = \mathbf{b}$. Then $a_i = b_i$ for all values of i, and so each term $(b_i - a_i)^2$ inside the square root of the Euclidean distance function

$$d^{(n)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2}$$

is 0, and hence $d^{(n)}(\mathbf{a},\mathbf{b}) = 0$. Conversely, since each term $(b_i - a_i)^2$ in $d^{(n)}(\mathbf{a},\mathbf{b})$ is non-negative, when $d^{(n)}(\mathbf{a},\mathbf{b}) = 0$ each term must be equal to 0, and so $a_i = b_i$ for all values of i. Thus $\mathbf{a} = \mathbf{b}$. Therefore, the distance $d^{(n)}(\mathbf{a},\mathbf{b})$ between two points is 0 if and only if $\mathbf{a} = \mathbf{b}$.

The next property is that the distance from $\mathbf{a}$ to $\mathbf{b}$ is the same as the distance from $\mathbf{b}$ to $\mathbf{a}$.

(M2) For all $\mathbf{a},\mathbf{b} \in \mathbb{R}^n$, $d^{(n)}(\mathbf{a},\mathbf{b}) = d^{(n)}(\mathbf{b},\mathbf{a})$.

This is referred to as the *symmetry property.*

This property is true since

$$d^{(n)}(\mathbf{a},\mathbf{b}) = \sqrt{(b_1-a_1)^2 + (b_2-a_2)^2 + \cdots + (b_n-a_n)^2},$$

$$d^{(n)}(\mathbf{b},\mathbf{a}) = \sqrt{(a_1-b_1)^2 + (a_2-b_2)^2 + \cdots + (a_n-b_n)^2}.$$

In the two expressions, corresponding terms inside the square root are equal, since $(p-q)^2 = (q-p)^2$ for any $p, q \in \mathbb{R}$. Hence $d^{(n)}(\mathbf{a},\mathbf{b}) = d^{(n)}(\mathbf{b},\mathbf{a})$.

The symmetry property of $d^{(n)}(\mathbf{a},\mathbf{b})$ allows us to use the following two forms interchangeably:

$$d^{(n)}(\mathbf{a},\mathbf{b}) = \sqrt{(b_1-a_1)^2 + (b_2-a_2)^2 + \cdots + (b_n-a_n)^2};$$

$$d^{(n)}(\mathbf{a},\mathbf{b}) = \sqrt{(a_1-b_1)^2 + (a_2-b_2)^2 + \cdots + (a_n-b_n)^2}.$$

The third property is the **Triangle Inequality**.

(M3) For all $\mathbf{a},\mathbf{b},\mathbf{c} \in \mathbb{R}^n$, $d^{(n)}(\mathbf{a},\mathbf{c}) \leq d^{(n)}(\mathbf{a},\mathbf{b}) + d^{(n)}(\mathbf{b},\mathbf{c})$.

There are many proofs of the Triangle Inequality, including geometrical ones. Most of the proofs that depend on geometry involve ideas outside the scope of this course, and so we do not give one here. In Subsection 5.4 we give a proof which is purely algebraic.

We summarize these results in the following theorem.

Theorem 5.1

The Euclidean distance function $d^{(n)}$ on $\mathbb{R}^n$ has the following properties.

For all $\mathbf{a},\mathbf{b},\mathbf{c} \in \mathbb{R}^n$:

(M1) $d^{(n)}(\mathbf{a},\mathbf{b}) \geq 0$, with equality holding if and only if $\mathbf{a} = \mathbf{b}$; — non-negativity

(M2) $d^{(n)}(\mathbf{a},\mathbf{b}) = d^{(n)}(\mathbf{b},\mathbf{a})$; — symmetry

(M3) $d^{(n)}(\mathbf{a},\mathbf{c}) \leq d^{(n)}(\mathbf{a},\mathbf{b}) + d^{(n)}(\mathbf{b},\mathbf{c})$. — Triangle Inequality

Remark

In *Unit 2*, we use these three properties as the definition for a general distance function, so you need to be familiar with them. In particular, be sure that you understand the meaning of the Triangle Inequality, as it is of basic importance.

The Triangle Inequality for $d^{(n)}$, like that for distances in $\mathbb{R}$, can be given in reversed form.

Theorem 5.2 Reverse Triangle Inequality

The Euclidean distance function $d^{(n)}$ on $\mathbb{R}^n$ has the following property.

(M3a) For all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$, $d^{(n)}(\mathbf{b}, \mathbf{c}) \geq \left| d^{(n)}(\mathbf{a}, \mathbf{c}) - d^{(n)}(\mathbf{a}, \mathbf{b}) \right|$.

Proof From (M3), we obtain, for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$,

$$d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c}).$$

Subtracting $d^{(n)}(\mathbf{a}, \mathbf{b})$ from each side we obtain

$$d^{(n)}(\mathbf{a}, \mathbf{c}) - d^{(n)}(\mathbf{a}, \mathbf{b}) \leq d^{(n)}(\mathbf{b}, \mathbf{c}). \tag{5.1}$$

Since $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are arbitrary points in $\mathbb{R}^2$, we can freely interchange their roles in the Triangle Inequality. Thus, interchanging the roles of $\mathbf{b}$ and $\mathbf{c}$ in (5.1), we get

$$d^{(n)}(\mathbf{a}, \mathbf{b}) - d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{c}, \mathbf{b}) = d^{(n)}(\mathbf{b}, \mathbf{c}),$$

by (M2). This can be rearranged to give

$$-d^{(n)}(\mathbf{b}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{c}) - d^{(n)}(\mathbf{a}, \mathbf{b}). \tag{5.2}$$

Combining inequalities (5.1) and (5.2) we obtain

$$-d^{(n)}(\mathbf{b}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{c}) - d^{(n)}(\mathbf{a}, \mathbf{b}) \leq d^{(n)}(\mathbf{b}, \mathbf{c}),$$

which is equivalent to the required result. ■

The Euclidean distance formula

$$d^{(n)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2}$$

defines a function of two variables $\mathbf{a}$ and $\mathbf{b}$, each from the set $\mathbb{R}^n$, which produce a real number, the distance. What are the domain and codomain of this function? To answer this question we need the idea of a *product set*.

Definition

Let $A_1, A_2, \ldots, A_k$ be sets. The **product set** $A_1 \times A_2 \times \cdots \times A_k$ is defined as the set of all ordered k-tuples $(a_1, a_2, \ldots, a_k)$ where $a_1 \in A_1, a_2 \in A_2, \ldots, a_k \in A_k$.

The product set is sometimes called the *Cartesian product*.

Remarks

(i) The term *product* is used for this set because, when $A_1, A_2, \ldots, A_k$ are finite sets with $m_1, m_2, \ldots, m_k$ elements respectively, the set $A_1 \times A_2 \times \cdots \times A_k$ has $m_1 m_2 \cdots m_k$ elements. For example, if $k = 2$, $A_1 = \{1, 2\}$ and $A_2 = \{1, 4, 9\}$, then

$$A_1 \times A_2 = \{(1,1), (1,4), (1,9), (2,1), (2,4), (2,9)\}$$

which has $2 \times 3 = 6$ elements.

(ii) If $k = 2$ and $A_1 = A_2 = \mathbb{R}$, then $A_1 \times A_2 = \mathbb{R} \times \mathbb{R}$, the set of all ordered pairs of real numbers. Thus $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$.
Similarly, if $A_1, A_2, \ldots, A_k$ are all $\mathbb{R}$, then the product set $A_1 \times A_2 \times \cdots \times A_k = \mathbb{R}^k$.

(iii) The definition can be extended to the product of infinitely many sets. In particular, the product of infinitely many copies of $\mathbb{R}$ gives the set $\mathbb{R}^\infty$, the set of all ordered infinite tuples of real numbers, of the form $(a_1, a_2, a_3, \ldots)$. This is equivalent to the set of all real sequences. We

This explains why we prefer (a_n) to $\{a_n\}$ as a notation for sequences, since parentheses imply an ordering whereas braces do not.

can write these infinite tuples in the form $(a_n)_{n=1}^{\infty}$, which coincides with the notation for sequences introduced in Subsection 1.3.

(iv) We often denote an element of a product set $A_1 \times A_2 \times \cdots \times A_k$ as $\mathbf{a} = (a_1, a_2, \ldots, a_k)$ and we say that two elements $\mathbf{a} = (a_1, a_2, \ldots, a_k)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_k)$ are equal if $a_i = b_i$ for $i = 1, 2, \ldots, k$.

Now, the domain of the distance function $d^{(1)}$ is simply the set of all ordered pairs (a, b) where $a \in \mathbb{R}$ and $b \in \mathbb{R}$ — that is, the domain is the product set $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. However, each of the arguments $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ of the distance function $d^{(2)}$ is itself an ordered pair of real numbers from $\mathbb{R}^2$. Thus the domain of the distance function $d^{(2)}$ is the set of all ordered pairs of elements of $\mathbb{R}^2$, i.e. the product set $\mathbb{R}^2 \times \mathbb{R}^2$. In the general case, the domain of the distance function $d^{(n)}$ is the product set $\mathbb{R}^n \times \mathbb{R}^n$.

For all n, the codomain of $d^{(n)}$ is $\mathbb{R}$.

Hence $d^{(n)}$ is a function

$$d^{(n)} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

from the set of all ordered pairs of ordered n-tuples of real numbers to the set of real numbers.

5.3 Continuity for functions between Euclidean spaces

In Section 3 we defined ε–δ continuity for functions from $\mathbb{R}$ to $\mathbb{R}$, and in Section 4 we extended this to continuity for functions from $\mathbb{R}^2$ to $\mathbb{R}$. To do this we had to alter the distance function from $d^{(1)}(a, b) = |b - a|$ to $d^{(2)}(\mathbf{a}, \mathbf{b})$. Now we wish to extend the definition of continuity to higher dimensions. It is a simple extension to define continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}$, as we merely change the distance function in the definition from $d^{(2)}(\mathbf{a}, \mathbf{b})$ to $d^{(n)}(\mathbf{a}, \mathbf{b})$. But we wish to define continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, for *any* positive integers n and m.

Look again at the definition for continuity for functions from $\mathbb{R}^2$ to $\mathbb{R}$. We have boxed the elements that depend on the choice of domain *and* codomain. Those involving the codomain have double boxes.

Let $A \subseteq \boxed{\mathbb{R}^2}$. A function $f \colon A \to \boxed{\boxed{\mathbb{R}}}$ is **continuous** at $\mathbf{a} \in A$ if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that, for all $\mathbf{x} \in A$,

$$\boxed{\boxed{|f(\mathbf{x}) - f(\mathbf{a})| < \varepsilon}} \quad \text{whenever} \quad \boxed{d^{(2)}(\mathbf{x}, \mathbf{a}) < \delta}.$$

We already know how to change the items in the single boxes — we simply replace $\mathbb{R}^2$ by $\mathbb{R}^n$ and $d^{(2)}(\mathbf{x}, \mathbf{a})$ by $d^{(n)}(\mathbf{x}, \mathbf{a})$. For the first double box there is no problem either: the codomain changes from $\mathbb{R}$ to $\mathbb{R}^m$. We also need to change $|f(\mathbf{x}) - f(\mathbf{a})|$ appropriately.

The insight needed here is that, for functions from $\mathbb{R}^2$ to $\mathbb{R}$, the modulus $|f(\mathbf{x}) - f(\mathbf{a})|$ *acts as the distance function in the codomain* $\mathbb{R}$. There, the values of $f(\mathbf{x})$ belong to the codomain $\mathbb{R}$; but, in functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, the values of $f(\mathbf{x})$ belong to $\mathbb{R}^m$. So we need to replace $|f(\mathbf{x}) - f(\mathbf{a})|$ by the Euclidean distance function $d^{(m)}(f(\mathbf{x}), f(\mathbf{a}))$ for the codomain $\mathbb{R}^m$.

> ***Definition***
>
> Let $A \subseteq \mathbb{R}^n$. A function $f\colon A \to \mathbb{R}^m$ is **continuous** at $\mathbf{a} \in A$ if, for each $\varepsilon > 0$, there exists a $\delta > 0$ such that, for all $\mathbf{x} \in A$,
>
> $$d^{(m)}(f(\mathbf{x}), f(\mathbf{a})) < \varepsilon \quad \text{whenever} \quad d^{(n)}(\mathbf{x}, \mathbf{a}) < \delta.$$

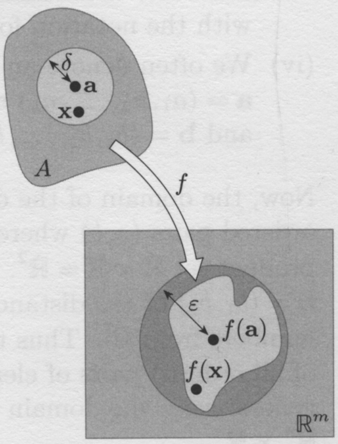


Figure 5.1

Remark

When $n = 1$ and $m = 1$, this definition reduces to the ε–δ definition of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ given in Section 3. Similarly, when $n = 2$ and $m = 1$, we obtain the definition of continuity for functions from $\mathbb{R}^2$ to $\mathbb{R}$ given in Section 4.

> ***Definition***
>
> Let $S \subseteq A \subseteq \mathbb{R}^n$ and let $f\colon A \to \mathbb{R}^m$ be a function. We say that f is **continuous** on S when f is continuous at each point in S.

We now consider some examples of continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, for various values of n and m.

Worked problem 5.1

Prove that the function $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$f(x_1, x_2, x_3) = (x_1 + x_3, x_2)$$

is continuous on $\mathbb{R}^3$.

Solution

Let $\mathbf{a}$ be a general point of $\mathbb{R}^3$ and let $\varepsilon > 0$ be given.

First we find an upper bound for $d^{(2)}(f(\mathbf{x}), f(\mathbf{a}))$ in terms of $d^{(3)}(\mathbf{x}, \mathbf{a})$.

We have, for $\mathbf{x} \in \mathbb{R}^3$,

$$\begin{aligned} d^{(2)}(f(\mathbf{x}), f(\mathbf{a})) &= \sqrt{((x_1 + x_3) - (a_1 + a_3))^2 + (x_2 - a_2)^2} \\ &= \sqrt{((x_1 - a_1) + (x_3 - a_3))^2 + (x_2 - a_2)^2} \\ &= \sqrt{(x_1 - a_1)^2 + 2(x_1 - a_1)(x_3 - a_3) + (x_3 - a_3)^2 + (x_2 - a_2)^2} \\ &= \sqrt{d^{(3)}(\mathbf{x}, \mathbf{a})^2 + 2(x_1 - a_1)(x_3 - a_3)}. \end{aligned}$$

Therefore

$$\left(d^{(2)}(f(\mathbf{x}), f(\mathbf{a}))\right)^2 = d^{(3)}(\mathbf{x}, \mathbf{a})^2 + 2(x_1 - a_1)(x_3 - a_3).$$

Now

$$\begin{aligned} 2(x_1 - a_1)(x_3 - a_3) &\le (x_1 - a_1)^2 + (x_3 - a_3)^2 \qquad (5.3) \\ &\le (x_1 - a_1)^2 + (x_2 - a_2)^2 + (x_3 - a_3)^2 \\ &= d^{(3)}(\mathbf{x}, \mathbf{a})^2. \end{aligned}$$

(5.3) is obtained from the inequality $(p - q)^2 \ge 0$, which gives $p^2 + q^2 - 2pq \ge 0$ or equivalently $2pq \le p^2 + q^2$.

Thus, $\left(d^{(2)}(f(\mathbf{x}), f(\mathbf{a}))\right)^2 \le 2\left(d^{(3)}(\mathbf{x}, \mathbf{a})\right)^2$, so that

$$d^{(2)}(f(\mathbf{x}), f(\mathbf{a})) \le \sqrt{2}\, d^{(3)}(\mathbf{x}, \mathbf{a}).$$

We now let $\delta = \varepsilon/\sqrt{2}$ and thus

$$d^{(2)}(f(\mathbf{x}), f(\mathbf{a})) < \varepsilon \quad \text{whenever} \quad d^{(3)}(\mathbf{x}, \mathbf{a}) < \delta.$$

We conclude that f is continuous at $\mathbf{a}$.

Since $\mathbf{a}$ is an arbitrary point of $\mathbb{R}^3$, f is continuous on $\mathbb{R}^3$. ■

Problem 5.3

Let $\mathbf{b} \in \mathbb{R}^m$ be fixed. Show that the constant function $f: \mathbb{R}^n \to \mathbb{R}^m$ given by

$$f(\mathbf{x}) = \mathbf{b}$$

is continuous on $\mathbb{R}^n$.

In the next worked problem we apply the definition of continuity to the distance function $d^{(n)}(\mathbf{x}, \mathbf{u})$ that maps a point in $\mathbb{R}^n$ onto its distance from a fixed point $\mathbf{u}$; it is thus a function from $\mathbb{R}^n$ to $\mathbb{R}$.

Worked problem 5.2

Let $\mathbf{u}$ be a fixed point in $\mathbb{R}^n$. Prove that the function $g: \mathbb{R}^n \to \mathbb{R}$ given by

$$g(\mathbf{x}) = d^{(n)}(\mathbf{x}, \mathbf{u})$$

is continuous on $\mathbb{R}^n$.

Solution

Let $\mathbf{a}$ be a general point of $\mathbb{R}^n$ and let $\varepsilon > 0$ be given.

In this solution we need to keep a clear head since the distance function is used in the domain and also as the function in the codomain.

First we find an upper bound for $d^{(1)}(g(\mathbf{x}), g(\mathbf{a})) = |g(\mathbf{x}) - g(\mathbf{a})|$ in terms of $d^{(n)}(\mathbf{x}, \mathbf{a})$.

Now, for $\mathbf{x} \in \mathbb{R}^n$,

$$|g(\mathbf{x}) - g(\mathbf{a})| = |d^{(n)}(\mathbf{x}, \mathbf{u}) - d^{(n)}(\mathbf{a}, \mathbf{u})|.$$

From the Reverse Triangle Inequality, we see that

$$|\, d^{(n)}(\mathbf{x}, \mathbf{u}) - d^{(n)}(\mathbf{a}, \mathbf{u})| \le d^{(n)}(\mathbf{x}, \mathbf{a}).$$

So, $|g(\mathbf{x}) - g(\mathbf{a})| \le d^{(n)}(\mathbf{x}, \mathbf{a})$.

Let $\delta = \varepsilon$. Then, whenever $d^{(n)}(\mathbf{x}, \mathbf{a}) < \delta$,

$$|g(\mathbf{x}) - g(\mathbf{a})| \le d^{(n)}(\mathbf{x}, \mathbf{a}) < \delta = \varepsilon.$$

Thus g is continuous at $\mathbf{a}$.

Since $\mathbf{a}$ is an arbitrary point of $\mathbb{R}^n$, g is continuous on $\mathbb{R}^n$. ■

As with the functions discussed in Sections 2, 3 and 4, there are Combination Rules for continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ and these often make it easier to show that a given function is continuous.

Combination Rules for continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

Let $A \subseteq \mathbb{R}^n$ and let $f: A \to \mathbb{R}^m$ and $g: A \to \mathbb{R}^m$ be continuous on A.

Then the following functions are continuous on A:

Sum $f + g: A \to \mathbb{R}^m$, defined by $(f+g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$;

Multiple $\lambda f: A \to \mathbb{R}^m$ where $\lambda \in \mathbb{R}$, defined by $(\lambda f)(\mathbf{x}) = \lambda \times f(\mathbf{x})$.

We ask you to prove these rules in the problems for this unit.

We do not give a Product Rule for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ because we have not defined a product for points of $\mathbb{R}^m$.

There is also a Restriction Rule for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ which allows us to restrict the domain of a continuous function to a smaller set and still have a continuous function. We state it without proof.

> ***Theorem 5.3 Restriction Rule***
>
> Let $A \subseteq \mathbb{R}^n$, let $f: A \to \mathbb{R}^m$ be continuous on A and let $B \subseteq A$. Then the restricted function $f|_B : B \to \mathbb{R}^m$ is continuous on B.

We can also generalize the Composition Rule. If the first function is from $\mathbb{R}^n$ to $\mathbb{R}^m$, then the second function must have its domain in $\mathbb{R}^m$, but may have any dimension for its codomain. Again we state it without proof.

> ***Theorem 5.4 Composition Rule***
>
> Let $A \subseteq \mathbb{R}^n$ and let $f: A \to \mathbb{R}^m$ be continuous on A. Let $f(A) \subseteq B \subseteq \mathbb{R}^m$ and let $g: B \to \mathbb{R}^k$ be continuous on B. Then the composed function $g \circ f: A \to \mathbb{R}^k$ is continuous on A.

Problem 5.4

Without using ε–δ techniques, prove that the function $g: \mathbb{R}^n \to \mathbb{R}$ given by

$$g(\mathbf{x}) = \left(d^{(n)}(\mathbf{x}, \mathbf{0})\right)^2$$

is continuous on $\mathbb{R}^n$.

This function maps $\mathbf{x}$ to the square of its distance from $\mathbf{0}$.

The Composition Rule is most useful in combination with projection functions from $\mathbb{R}^n$ to $\mathbb{R}$, which are simple extensions of the projection functions from $\mathbb{R}^2$ to $\mathbb{R}$ introduced in Subsection 4.2.

> ***Definition***
>
> The n functions $p_j : \mathbb{R}^n \to \mathbb{R}$ given by
>
> $$p_j(x_1, x_2, \ldots, x_n) = x_j, \quad \text{for } j = 1, 2, \ldots, n,$$
>
> are known as **projection functions**.

Remarks

(i) Each of these functions maps a point $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ onto one of its coordinates. Thus $p_1(\mathbf{x}) = x_1$, $p_2(\mathbf{x}) = x_2$, and so on.

(ii) By imitating the proof of Worked problem 4.1, we can show that the projection functions are continuous on $\mathbb{R}^n$.

We ask you to prove this in the problems for this unit.

Note that the effect of composing each separate p_j with a function f is to produce the separate function values of f. For example, composing p_1, p_2 and p_3 with the function

$$h: \mathbb{R}^5 \to \mathbb{R}^3, \quad h(x_1, x_2, x_3, x_4, x_5) = (2x_1, x_2^2 + 3x_4, x_3 + x_5),$$

gives

$$(p_1 \circ h)(x_1, x_2, x_3, x_4, x_5) = 2x_1,$$
$$(p_2 \circ h)(x_1, x_2, x_3, x_4, x_5) = x_2^2 + 3x_4,$$
$$(p_3 \circ h)(x_1, x_2, x_3, x_4, x_5) = x_3 + x_5.$$

The Composition Rule and the projection functions $p_j : \mathbb{R}^m \to \mathbb{R}$ give us a useful technique for proving that a function f from $\mathbb{R}^n$ to $\mathbb{R}^m$ is continuous. The main idea is that, instead of having to prove directly that f is continuous, we form the composite of each of the projection functions p_j with the function f, and then prove that each of the functions $p_j \circ f$ is continuous. Since the functions $p_j \circ f$ are functions from $\mathbb{R}^m$ to $\mathbb{R}$, this is usually much simpler than proving directly that f is continuous.

The following theorem, which enables us to use projection functions in this way, exploits the fact that the Euclidean distance function $d^{(m)}(f(\mathbf{x}), f(\mathbf{a}))$ can be expressed in terms of the functions $p_j \circ f$. To see this, let $f: A \to \mathbb{R}^m$, where $A \subseteq \mathbb{R}^n$. Then the value of f at a point $\mathbf{x} \in A$ can be expressed as

$$f(\mathbf{x}) = ((p_1 \circ f)(\mathbf{x}), \ldots, (p_m \circ f)(\mathbf{x})).$$

Thus, if $\mathbf{a} \in A$, then

$$d^{(m)}(f(\mathbf{x}), f(\mathbf{a})) = \sqrt{\Big((p_1 \circ f)(\mathbf{x}) - (p_1 \circ f)(\mathbf{a})\Big)^2 + \cdots + \Big((p_m \circ f)(\mathbf{x}) - (p_m \circ f)(\mathbf{a})\Big)^2}.$$

If we know that $p_1 \circ f, \ldots, p_m \circ f$ are all continuous at $\mathbf{a}$, then this equality enables us to deduce that f is continuous at $\mathbf{a}$.

> ***Theorem 5.5***
>
> Let $A \subseteq \mathbb{R}^n$ and let $f: A \to \mathbb{R}^m$. The function f is continuous at $\mathbf{a} \in A$ if and only if $p_j \circ f: A \to \mathbb{R}$ is continuous at $\mathbf{a}$, for $j = 1, 2, \ldots, m$, where each $p_j : \mathbb{R}^m \to \mathbb{R}$ is a projection function.

Proof The proof is in two parts. In each part we deal with the m functions $p_1 \circ f, p_2 \circ f, \ldots, p_m \circ f$ simultaneously.

First, we prove that if f is continuous at $\mathbf{a}$ then the functions $p_j \circ f$ are also continuous at $\mathbf{a}$. This follows directly from the Composition Rule. Since f is continuous at $\mathbf{a}$ and all of the p_j are continuous on $\mathbb{R}^m$, the functions $p_j \circ f$ are also continuous at $\mathbf{a}$, for $j = 1, 2, \ldots, m$.

Conversely, we now prove that if the functions $p_j \circ f$ are continuous at $\mathbf{a}$ then f is also continuous at $\mathbf{a}$ — that is, we show that, given any $\varepsilon > 0$, there is $\delta > 0$ such that

This is the useful part of Theorem 5.5.

$$d^{(m)}(f(\mathbf{x}), f(\mathbf{a})) < \varepsilon \quad \text{whenever} \quad d^{(n)}(\mathbf{x}, \mathbf{a}) < \delta.$$

Let $\varepsilon > 0$ be given. Since the functions $p_j \circ f$ are continuous at $\mathbf{a}$, we can find positive numbers $\delta_1, \delta_2, \ldots, \delta_m$, such that

We find a separate δ for each projection function.

$$\begin{aligned}
&\text{whenever } d^{(n)}(\mathbf{x}, \mathbf{a}) < \delta_1, \text{ then } |(p_1 \circ f)(\mathbf{x}) - (p_1 \circ f)(\mathbf{a})| < \varepsilon/\sqrt{m},\\
&\text{whenever } d^{(n)}(\mathbf{x}, \mathbf{a}) < \delta_2, \text{ then } |(p_2 \circ f)(\mathbf{x}) - (p_2 \circ f)(\mathbf{a})| < \varepsilon/\sqrt{m},\\
&\vdots\\
&\text{whenever } d^{(n)}(\mathbf{x}, \mathbf{a}) < \delta_m, \text{ then } |(p_m \circ f)(\mathbf{x}) - (p_m \circ f)(\mathbf{a})| < \varepsilon/\sqrt{m}.
\end{aligned}$$

See the remark below concerning $\varepsilon/\sqrt{m}$.

We require $d^{(n)}(\mathbf{x}, \mathbf{a})$ to be less than each δ_j, and so we set $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_m\}$. Then, whenever $d^{(n)}(\mathbf{x}, \mathbf{a}) < \delta$,

$$\begin{aligned}
&d^{(m)}(f(\mathbf{x}), f(\mathbf{a}))\\
&\quad = \sqrt{\Big((p_1 \circ f)(\mathbf{x}) - (p_1 \circ f)(\mathbf{a})\Big)^2 + \cdots + \Big((p_m \circ f)(\mathbf{x}) - (p_m \circ f)(\mathbf{a})\Big)^2}\\
&\quad < \sqrt{(\varepsilon/\sqrt{m})^2 + \cdots + (\varepsilon/\sqrt{m})^2} = \sqrt{m(\varepsilon/\sqrt{m})^2} = \sqrt{m\varepsilon^2/m} = \varepsilon.
\end{aligned}$$

Thus f is continuous at $\mathbf{a}$. ■

Remark

The appearance of the expression $\varepsilon/\sqrt{m}$, rather than simply ε, may surprise you. In fact, the division by $\sqrt{m}$ is there to enable us to end up with $d^{(m)}(f(\mathbf{x}), f(\mathbf{a})) < \varepsilon$. Since m is fixed, $\varepsilon/\sqrt{m}$ serves as well as ε in the condition that the functions $p_j \circ f$ are continuous at $\mathbf{a}$.

Worked problem 5.3

Let $A = [0,1] \times [-0.1, 0.1] = \{(x_1, x_2): 0 \le x_1 \le 1, -0.1 \le x_2 \le 0.1\}$ and define $f: A \to \mathbb{R}^3$ by

$$f(x_1, x_2) = ((1 + x_2 \cos \pi x_1) \cos 2\pi x_1, (1 + x_2 \cos \pi x_1) \sin 2\pi x_1, x_2 \sin \pi x_1).$$

Prove that f is continuous on A.

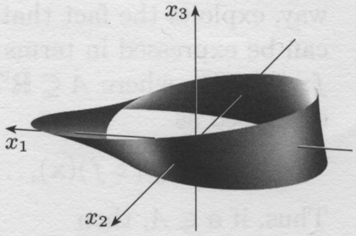


Figure 5.2 The image set $f(A)$ in Worked problem 5.3 is a Möbius band

Solution

Consider the composition of f with the three projection functions p_1, p_2 and p_3 from $\mathbb{R}^3$ to $\mathbb{R}$. These are given by

$$(p_1 \circ f)(x_1, x_2) = (1 + x_2 \cos \pi x_1) \cos 2\pi x_1,$$

$$(p_2 \circ f)(x_1, x_2) = (1 + x_2 \cos \pi x_1) \sin 2\pi x_1,$$

$$(p_3 \circ f)(x_1, x_2) = x_2 \sin \pi x_1.$$

If we can show that each $p_j \circ f$ is continuous on A then it follows from Theorem 5.5 that f is continuous on A. We prove that $p_1 \circ f$ is continuous on A; the proofs for $p_2 \circ f$ and $p_3 \circ f$ are similar.

To show that $(p_1 \circ f)(x_1, x_2) = (1 + x_2 \cos \pi x_1) \cos 2\pi x_1$ is continuous on A, we show that the constituent parts are continuous and then combine them using the Combination Rules.

We first verify that the functions from $\mathbb{R}^2$ to $\mathbb{R}$ given by $(x_1, x_2) \mapsto 2\pi x_1$, $(x_1, x_2) \mapsto \pi x_1$, $(x_1, x_2) \mapsto x_2$ are continuous. We use the result of Worked problem 4.3, where we proved the continuity of the function $g: \mathbb{R}^2 \to \mathbb{R}$ given by $g(\mathbf{x}) = u_1 x_1 + u_2 x_2$, where u_1 and u_2 are fixed. If we set $u_1 = 2\pi$ and $u_2 = 0$ we can conclude that the function $(x_1, x_2) \mapsto 2\pi x_1 + 0x_2 = 2\pi x_1$ is continuous. Similarly, by taking the values of (u_1, u_2) to be $(\pi, 0)$ and $(0, 1)$ respectively, we can conclude that $(x_1, x_2) \mapsto \pi x_1$ and $(x_1, x_2) \mapsto x_2$ are continuous.

We are also restricting the functions to the set A, and so we need to apply the Restriction Rule for functions from $\mathbb{R}^2$ to $\mathbb{R}$ to these three functions. Then, since $x \mapsto \cos x$ is a basic continuous function from $\mathbb{R}$ to $\mathbb{R}$, we can apply the Composition Rule. This shows that $(x_1, x_2) \mapsto \cos 2\pi x_1$ and $(x_1, x_2) \mapsto \cos \pi x_1$ are continuous. We can now use the Combination Rules for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$. We use the Product Rule twice to deduce that $(x_1, x_2) \mapsto x_2 \cos \pi x_1 \cos 2\pi x_1$ is continuous and then use the Sum Rule to conclude that $(x_1, x_2) \mapsto \cos 2\pi x_1 + x_2 \cos \pi x_2 \cos 2\pi x_1$ is continuous. ■

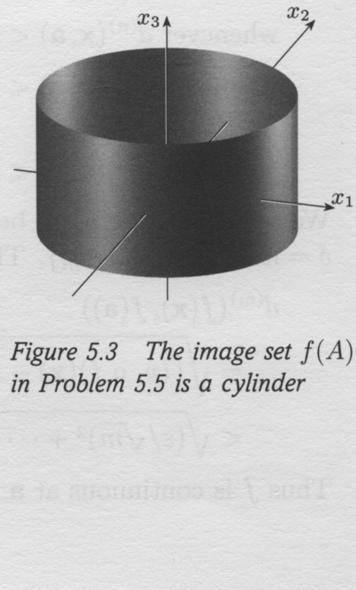


Figure 5.3 The image set $f(A)$ in Problem 5.5 is a cylinder

Problem 5.5

Let $A = [0,1] \times [-1, 1] = \{(x_1, x_2): 0 \le x_1 \le 1, -1 \le x_2 \le 1\}$ and define $f: A \to \mathbb{R}^3$ by

$$f(x_1, x_2) = (\cos 2\pi x_1, \sin 2\pi x_1, x_2).$$

Prove that f is continuous on A.

For completeness, we conclude this subsection with the definition of discontinuity for a function $f: \mathbb{R}^n \to \mathbb{R}^m$.

> ***Definition***
>
> Let $S \subseteq A \subseteq \mathbb{R}^n$ and let $f: A \to \mathbb{R}^m$ be a function.
>
> Then f is **discontinuous** at $\mathbf{a} \in A$ if f is not continuous at $\mathbf{a}$.
>
> Furthermore, f is **discontinuous** on S if f is discontinuous at at least one point in S.

5.4 Proof of the Triangle Inequality

In this subsection we give a proof of the Triangle Inequality

This subsection is not assessed.

$$d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c})$$

using only algebraic methods. The proof makes use of another inequality, the *Cauchy–Schwarz Inequality*, that has important applications in many areas of mathematics. We first state and prove this inequality.

> ***Theorem 5.6 Cauchy–Schwarz Inequality***
>
> Let $(r_1, r_2, \ldots, r_n)$ and $(s_1, s_2, \ldots, s_n)$ be points in $\mathbb{R}^n$. Then
>
> $$\left(\sum_{j=1}^{n} r_j s_j\right)^2 \leq \sum_{j=1}^{n} r_j^2 \sum_{j=1}^{n} s_j^2. \qquad (5.4)$$

Remark

A useful technique for understanding a general statement is to specialize to small particular values.

When $n = 1$, the Cauchy–Schwarz Inequality becomes

$$(r_1 s_1)^2 \leq (r_1^2)(s_1^2).$$

Since the two sides are equal, this is clearly true.

When $n = 2$, the inequality becomes

$$(r_1 s_1 + r_2 s_2)^2 \leq (r_1^2 + r_2^2)(s_1^2 + s_2^2).$$

Multiplying out, we obtain

$$r_1^2 s_1^2 + 2r_1 r_2 s_1 s_2 + r_2^2 s_2^2 \leq r_1^2 s_1^2 + r_1^2 s_2^2 + r_2^2 s_1^2 + r_2^2 s_2^2$$

and this inequality is true if and only if

$$0 \leq r_1^2 s_2^2 - 2r_1 r_2 s_1 s_2 + r_2^2 s_1^2 = (r_1 s_2 - r_2 s_1)^2.$$

But this is always true as the right-hand side is non-negative. Thus the Cauchy–Schwarz Inequality holds for $n = 2$.

We could give a proof of the Cauchy–Schwarz Inequality by developing the ideas indicated in the remark for a general value of n. However, we give a different proof, that involves a strategy based on some properties of quadratic equations with which you may well be familiar.

Let $A, B, C \in \mathbb{R}$, where $A \neq 0$, and suppose that

$$Ax^2 + Bx + C \geq 0, \qquad \text{for all } x.$$

Then the quadratic equation $Ax^2 + Bx + C = 0$ has at most one real root. For, if the equation had two distinct real roots, then the graph of $f(x) = Ax^2 + Bx + C$ would cut the x-axis in two places and there would be a range of x for which $f(x) < 0$, i.e. $Ax^2 + Bx + C < 0$.

The roots of the quadratic equation are given by

$$x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}.$$

When there is at most one real root, $B^2 - 4AC \leq 0$, so that $B^2 \leq 4AC$. This is the idea we shall use.

The expression $B^2 - 4AC$ is called the *discriminant* of the quadratic expression.

Proof If $r_j = 0$ for $j = 1, 2, \ldots, n$, then inequality (5.4) becomes

$$0 \leq 0 \times \sum_{j=1}^{n} s_j^2 = 0,$$

which is clearly true. Therefore, we need only consider cases where $(r_1, r_2, \ldots, r_n) \neq (0, 0, \ldots, 0)$.

We now consider the expression

$$\sum_{j=1}^{n} (r_j x + s_j)^2.$$

Since a sum of squares is always non-negative, it follows that, for any real number x,

$$\sum_{j=1}^{n} (r_j x + s_j)^2 \geq 0,$$

with equality if and only if each term in the sum is 0.

Multiplying out the left-hand side of this inequality, we obtain

$$\sum_{j=1}^{n} (r_j^2 x^2 + 2 r_j s_j x + s_j^2) \geq 0,$$

or equivalently

$$\left(\sum_{j=1}^{n} r_j^2 \right) x^2 + 2 \left(\sum_{j=1}^{n} r_j s_j \right) x + \left(\sum_{j=1}^{n} s_j^2 \right) \geq 0.$$

The left-hand side is a quadratic expression $Ax^2 + Bx + C$ with

$$A = \sum_{j=1}^{n} r_j^2, \quad B = 2 \sum_{j=1}^{n} r_j s_j \quad \text{and} \quad C = \sum_{j=1}^{n} s_j^2.$$

A is not zero since $(r_1, r_2, \ldots, r_n) \neq (0, 0, \ldots, 0)$.

So we have, for all x,

$$Ax^2 + Bx + C \geq 0,$$

and hence

$$B^2 \leq 4AC.$$

Substituting for A, B and C in this inequality, we obtain

$$4 \left(\sum_{j=1}^{n} r_j s_j \right)^2 \leq 4 \left(\sum_{j=1}^{n} r_j^2 \right) \left(\sum_{j=1}^{n} s_j^2 \right).$$

Dividing each side by 4 gives the required inequality. ■

Remark

If we take the square root of both sides of (5.4), we obtain the inequality

$$\sum_{j=1}^{n} r_j s_j \le \left(\sum_{j=1}^{n} r_j^2\right)^{\frac{1}{2}} \left(\sum_{j=1}^{n} s_j^2\right)^{\frac{1}{2}}. \qquad (5.5)$$

This inequality is also known as the **Cauchy–Schwarz Inequality**.

We can now prove the Triangle Inequality (M3).

> ***Triangle Inequality***
>
> For all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$, $d^{(n)}(\mathbf{a}, \mathbf{c}) \le d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c})$.

In this proof, we write the Triangle Inequality in a form which enables us to see that it follows from the Cauchy–Schwarz Inequality. To do this we use the formula for the Euclidean distance function. There is nothing very deep here, but the algebraic manipulation requires some care, as it involves sums of n real numbers and their squares.

Proof Let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be in $\mathbb{R}^n$. For convenience, we introduce the abbreviations $\mathbf{r} = \mathbf{a} - \mathbf{b}$ and $\mathbf{s} = \mathbf{b} - \mathbf{c}$, so that $\mathbf{r} + \mathbf{s} = \mathbf{a} - \mathbf{c}$. Thus

$$d^{(n)}(\mathbf{a}, \mathbf{b}) = d^{(n)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(n)}(\mathbf{r}, \mathbf{0}) = \left(\sum_{j=1}^{n} r_j^2\right)^{\frac{1}{2}},$$

$$d^{(n)}(\mathbf{b}, \mathbf{c}) = d^{(n)}(\mathbf{b} - \mathbf{c}, \mathbf{0}) = d^{(n)}(\mathbf{s}, \mathbf{0}) = \left(\sum_{j=1}^{n} s_j^2\right)^{\frac{1}{2}},$$

$$d^{(n)}(\mathbf{a}, \mathbf{c}) = d^{(n)}(\mathbf{a} - \mathbf{c}, \mathbf{0}) = d^{(n)}(\mathbf{r} + \mathbf{s}, \mathbf{0}) = \left(\sum_{j=1}^{n} (r_j + s_j)^2\right)^{\frac{1}{2}}.$$

We are aiming to obtain an expression to which we can apply the Cauchy–Schwarz Inequality.

Squaring and expanding, we find that

$$\begin{aligned}\left(d^{(n)}(\mathbf{a}, \mathbf{c})\right)^2 &= \sum_{j=1}^{n} (r_j + s_j)^2 \\ &= \left(\sum_{j=1}^{n} r_j^2\right) + 2\left(\sum_{j=1}^{n} r_j s_j\right) + \left(\sum_{j=1}^{n} s_j^2\right).\end{aligned}$$

We square to avoid the complication of having to deal with the square root.

If we now use the Cauchy–Schwarz Inequality (5.5) to estimate the middle term on the right-hand side, we obtain

$$\begin{aligned}\left(d^{(n)}(\mathbf{a}, \mathbf{c})\right)^2 &\le \left(\sum_{j=1}^{n} r_j^2\right) + 2\left(\sum_{j=1}^{n} r_j^2\right)^{\frac{1}{2}} \left(\sum_{j=1}^{n} s_j^2\right)^{\frac{1}{2}} + \left(\sum_{j=1}^{n} s_j^2\right) \\ &= \left(\left(\sum_{j=1}^{n} r_j^2\right)^{\frac{1}{2}} + \left(\sum_{j=1}^{n} s_j^2\right)^{\frac{1}{2}}\right)^2 \\ &= \left(d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c})\right)^2.\end{aligned}$$

Taking square roots, we obtain

$$d^{(n)}(\mathbf{a}, \mathbf{c}) \le d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c}),$$

which completes the proof of the Triangle Inequality. ■

Solutions to problems

1.1 **(a)** This is a function from $\mathbb{Z}$ to $\mathbb{N}$, since the square of an integer is a non-negative integer and so adding 1 to this number always gives a natural number.

(b) This is not a function from $\mathbb{Z}$ to $\mathbb{N}$, since it associates with a given integer a *set* of natural numbers, and this is not an element of the codomain.

(c) This is not a function from $\mathbb{Z}$ to $\mathbb{N}$, since the integer 0 has no well-defined image.

1.2 The answers to (a) and (b) can be read off immediately from the box.

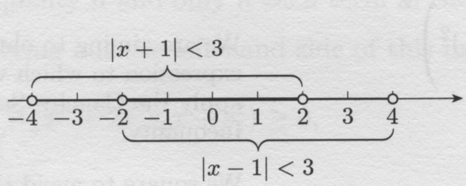
(a) The solution set is $(-3, 3)$.

(b) The solution set is $\varnothing$.

(c) $|3x| < 3 \iff |3|\,|x| < 3 \iff 3|x| < 3 \iff |x| < 1$. Thus $\{x \in \mathbb{R} : |3x| < 3\} = \{x \in \mathbb{R} : |x| < 1\}$, so the solution set is $(-1, 1)$.

1.3 **(a)** By the first result in the box, with $b = 1$ and $c = 3$, the solution set is the open interval $(1-3, 1+3) = (-2, 4)$.

(b) From (a), $|x-1| < 3$ gives x in $(-2, 4)$. Similarly, $|x+1| < 3$ if and only if x is in the open interval $(-4, 2)$. Since both inequalities must hold, we need the intersection of the intervals $(-2, 4)$ and $(-4, 2)$. This is the open interval $(-2, 2)$.



1.4 For $c \geq 0$, we have

$$|x| \leq c \quad \text{if and only if} \quad -c \leq x \leq c.$$

Put $x = |a|$ and $c = |b|$. Then,

$$||a|| = |a| \leq |b| \quad \text{if and only if} \quad -|b| \leq |a| \leq |b|.$$

1.5 **(a)** $|a+b| = |3+4| = |7| = 7$
and $|a| + |b| = |3| + |4| = 3 + 4 = 7$;
since $7 \leq 7$, the Triangle Inequality holds.

(b) $|a+b| = |-3+4| = |1| = 1$
and $|a| + |b| = |-3| + |4| = 3 + 4 = 7$;
since $1 \leq 7$, the Triangle Inequality holds.

(c) $|a+b| = |-3-4| = |-7| = 7$
and $|a| + |b| = |-3| + |-4| = 3 + 4 = 7$;
since $7 \leq 7$, the Triangle Inequality holds.

(Notice that equality holds in (a) and (c), while strict inequality holds in (b). Can you generalize this result?)

1.6 **(a)** We use the Triangle Inequality twice.
Let $d = b + c$. Then,

$$|a+b+c| = |a+d| \leq |a| + |d|,$$

by the Triangle Inequality. Also,

$$|d| = |b+c| \leq |b| + |c|.$$

Combining these two inequalities gives

$$|a+b+c| \leq |a| + |b| + |c|.$$

(b) The Triangle Inequality states that, for *all* real numbers x and y,

$$|x+y| \leq |x| + |y|.$$

Let $x = a$ and $y = -b$. Then,

$$|a-b| = |a + (-b)| \leq |a| + |-b| = |a| + |b|.$$

1.7 We first write the Triangle Inequality using p and q instead of a and b, to obtain $|p+q| \leq |p| + |q|$, where p and q are any real numbers. We now put $p = b - a$ and $q = c - b$, so that $p + q = c - a$, where a, b and c are any real numbers, and hence the inequality becomes

$$|c-a| \leq |b-a| + |c-b|.$$

Conversely, we rewrite the above inequality with a, b and c replaced by p, q and r, where p, q and r are any real numbers, giving $|r-p| \leq |q-p| + |r-q|$. We now put $a = q - p$ and $b = r - q$, so that $a + b = r - p$, where a and b are any real numbers, and hence the inequality becomes

$$|a+b| \leq |a| + |b|.$$

Thus the two inequalities are equivalent.

1.8 Let $\varepsilon > 0$ be given. Here $|a_n| = 1/\sqrt{n}$ and

$$1/\sqrt{n} < \varepsilon \quad \text{if and only if} \quad n > 1/\varepsilon^2.$$

If we set $N = [1/\varepsilon^2]$ then

$$|a_n| = 1/\sqrt{n} < \varepsilon \quad \text{whenever } n > N.$$

Hence (a_n) is null.

1.9 For all n, $|a_n| = |2| = 2$. We take any $\varepsilon \leq 2$, say $\varepsilon = 1$, and so, for all n, $|a_n| = 2 > 1 = \varepsilon$. We deduce that (a_n) is not a null sequence.

1.10 (a_n) is a null sequence and $\lambda \in \mathbb{R}$.
When $\lambda = 0$, we have $\lambda a_n = 0$ for all n and so (λa_n) is a null sequence (see Worked problem 1.3(a)).
When $\lambda \neq 0$, let $\varepsilon > 0$ be given. Then

$$|\lambda a_n| = |\lambda|\,|a_n| < \varepsilon \quad \text{if and only if} \quad |a_n| < \varepsilon/|\lambda|.$$

Since (a_n) is null, there is a natural number $N \in \mathbb{N}$ such that $|a_n| < \varepsilon/|\lambda|$ whenever $n > N$. Hence, $|\lambda a_n| < \varepsilon$ whenever $n > N$ and so (λa_n) is null.

(Notice that this proof uses the fact that the definition of a null sequence relies on having $|a_n| < \varepsilon$ for *all* $\varepsilon > 0$, and so in particular it applies to $\varepsilon/|\lambda|$ for any non-zero constant λ.)

1.11 Since (a_n) is null, the Multiple Rule implies that (λa_n) is null. Now $|b_n| \leq \lambda |a_n| = |\lambda a_n|$ (since $\lambda \geq 0$). It follows from the Squeeze Rule that (b_n) is null.

1.12 A sequence (a_n) is divergent if the sequence $(a_n - l)$ is not a null sequence for *any* value of l. That is, for any l, there is at least one $\varepsilon > 0$ for which there is *no* natural number N such that $|a_n - l| < \varepsilon$ whenever $n > N$. Alternatively, whatever the value of l, there exists an $\varepsilon > 0$ such that, no matter what choice of N we make, we can always find a natural number $n > N$ for which $|a_n - l| \geq \varepsilon$.

1.13 Let

$$a_n = \frac{4n^2 + 1000^n}{n! + 7n^3}.$$

Dividing both the numerator and denominator of a_n by the dominant term $n!$ gives

$$a_n = \frac{4(n^2/n!) + 1000^n/n!}{1 + 7(n^3/n!)}.$$

The sequences $(n^2/n!)$, $(1000^n/n!)$, $(n^3/n!)$ are all basic null sequences and any constant sequence converges to its constant value. Hence, by the Combination Rules for convergent sequences,

$$a_n \to \frac{(4 \times 0) + 0}{1 + (7 \times 0)} = 0 \quad \text{as } n \to \infty.$$

Thus (a_n) converges to 0.

1.14 Write $a_n = 2^n$ and let $n_k = k^2$ for all $k \in \mathbb{N}$. Then, (n_k) is a strictly increasing sequence of positive integers, and $a_{n_k} = 2^{k^2}$. Thus $(a_{n_k}) = (2^{k^2})$ is a subsequence of $(a_n) = (2^n)$.

1.15 Let

$$a_n = (-1)^n \frac{1-n}{n}.$$

Consider the even and odd subsequences (a_{2k}) and (a_{2k-1}).

For (a_{2k}), we have

$$a_{2k} = (-1)^{2k} \frac{1-2k}{2k} = \frac{1-2k}{2k}.$$

Dividing both the numerator and denominator by the dominant term $2k$ gives

$$a_{2k} = \frac{\frac{1}{2k} - 1}{1}.$$

Now $\left(\frac{1}{2k}\right) = \left(\frac{1}{2}\frac{1}{k}\right)$ is a null sequence (since $\left(\frac{1}{k}\right)$ is a basic null sequence), using the Multiple Rule for null sequences with $\lambda = \frac{1}{2}$. So the Combination Rules for convergent sequences imply that

$$a_{2k} \to -1 \quad \text{as } k \to \infty.$$

For (a_{2k-1}), we have

$$\begin{aligned} a_{2k-1} &= (-1)^{2k-1} \frac{1-(2k-1)}{2k-1} \\ &= (-1)\frac{2-2k}{2k-1} = \frac{2k-2}{2k-1}. \end{aligned}$$

Dividing both the numerator and denominator by the dominant term $2k$ gives

$$a_{2k-1} = \frac{1 - \frac{1}{k}}{1 - \frac{1}{2k}}.$$

Now $\left(\frac{1}{k}\right)$ and $\left(\frac{1}{2k}\right)$ are null sequences, so the Combination Rules for convergent sequences imply that

$$a_{2k-1} \to 1 \quad \text{as} \quad k \to \infty.$$

Since the even and odd subsequences have different limits, we deduce from Theorem 1.3 that (a_n) is divergent.

2.1 Let a be a general point in $\mathbb{R}$. We show that, if (x_n) is any sequence in $\mathbb{R}$ that converges to a, then the sequence $(f(x_n))$ converges to $f(a) = a^3 + 3$.

Let (x_n) be any sequence that converges to a. Now

$$f(x_n) = x_n^3 + 3 = (x_n \times x_n \times x_n) + 3.$$

Therefore, by the Product Rule for sequences (applied twice), the Sum Rule for sequences, and the fact that constant sequences converge to their constant value,

$$f(x_n) \to (a \times a \times a) + 3 = a^3 + 3 = f(a) \text{ as } n \to \infty.$$

Thus, $f(x_n) \to f(a)$ as $x_n \to a$, and so f is continuous at a. Since a is an arbitrary point of $\mathbb{R}$, f is continuous on $\mathbb{R}$.

2.2 There are two cases to consider: $a > 0$ and $a < 0$.

Let $a > 0$, so that $f_2(a) = 1$. Let (x_n) be any sequence in $\mathbb{R}$ such that $x_n \to a$ as $n \to \infty$. Then we can find N such that $x_n > 0$ for $n > N$, for example by taking N so that, for $n > N$, $|x_n - a| < |a|$ (so, the distance of x_n from a is less than the distance of a from 0). Thus, for $n > N$, $f_2(x_n) = f(a) = 1$, so $f_2(x_n) \to 1 = f_2(a)$ as $n \to \infty$. Hence f_2 is continuous at a.

For $a < 0$, we proceed in a similar manner. Noting that $f_2(a) = 0$, we let (x_n) be any sequence in $\mathbb{R}$ that converges to a. Then we can find N such that $x_n < 0$ for $n > N$, so that $f_2(x_n) = f_2(a) = 0$ for $n > N$. Hence $f_2(x_n) \to 0 = f_2(a)$ as $n \to \infty$, and f_2 is continuous at a.

2.3 We need to show that, for any null sequence (x_n), $f(x_n) \to f(0) = 0$.

So suppose (x_n) is null. Then, for each n, *either* $x_n \neq 0$ and

$$|f(x_n)| = |x_n^2 \sin(\pi/x_n)| = |x_n^2|\,|\sin(\pi/x_n)|,$$

or $x_n = 0$ and $f(x_n) = 0$.

Now $|\sin a| \leq 1$ for all real numbers a and so, when $x_n \neq 0$,

$$|f(x_n)| = |x_n^2|\,|\sin(\pi/x_n)| \leq |x_n^2|.$$

When $x_n = 0$,

$$|f(x_n)| = 0 \leq |x_n^2| = 0.$$

Therefore $|f(x_n)| \leq |x_n^2|$ for all n.

But, by the Product Rule for null sequences, (x_n^2) is a null sequence and so, by the Squeeze Rule, $f(x_n) \to 0$ as $n \to \infty$; that is, $f(x_n) \to f(0)$ as $x_n \to 0$. Thus f is continuous at 0.

2.4 The functions $x \mapsto e^x$, $x \mapsto -4x^3$ and $x \mapsto \sin x$ are basic continuous functions with domain $\mathbb{R}$. The function $x \mapsto 1/x$ is a basic continuous function with domain $\mathbb{R} - \{0\}$. Hence, by the Restriction Rule, $x \mapsto 1/x$ and $x \mapsto -4x^3$ are continuous on $(0, \infty)$. Also, by the Composition Rule, $x \mapsto \sin(1/x)$ and $x \mapsto e^{-4x^3}$ are continuous on $(0, \infty)$. It follows from the Multiple Rule that $x \mapsto 2\sin(1/x)$ and $x \mapsto -3e^{-4x^3}$ are continuous on $(0, \infty)$. Finally, by the Sum Rule, f is continuous on $(0, \infty)$.

2.5 We are given that f is continuous on $\mathbb{R}$. The function $x \mapsto |x|$ is a basic continuous function on $\mathbb{R}$. Hence, by the Composition Rule, the function $|f|$ is continuous on $\mathbb{R}$.

2.6 We first need to show that f is continuous on $[0, 3]$. The domain of the basic continuous function $x \mapsto \log x$ is $(0, \infty)$, the domain of the basic continuous function $x \mapsto 1 + x^2$ is $\mathbb{R}$ and the image set of $x \mapsto 1 + x^2$ is $[1, \infty)$. It follows (by the Composition Rule) that $x \mapsto \log(1 + x^2)$ is continuous on $\mathbb{R}$ and hence (by the Restriction Rule) on $[0, 3]$. Similarly, $x \mapsto 3\cos \pi x$ is a continuous function on $\mathbb{R}$ and hence on $[0, 3]$. So it follows (by the Sum Rule) that f is a continuous function on $[0, 3]$.

Since $f(0) = 3 > 0$ and $f(3) = \log(10) - 3 < 0$, it follows from the Intermediate Value Theorem that there is a number $c \in (0, 3)$ such that $f(c) = 0$.

2.7 First we show that $g: [0, 1] \to \mathbb{R}$ given by $g(x) = f(x) - x$ is continuous. Since $x \mapsto -x$ is a basic continuous function on $\mathbb{R}$, it is continuous on $[0, 1]$ (by the Restriction Rule). As f is continuous on $[0, 1]$, so is $g(x)$ (by the Sum Rule).

As $f(0) \geq 0$,

$$g(0) = f(0) - 0 \geq 0$$

and, as $f(1) \leq 1$,

$$g(1) = f(1) - 1 \leq 1 - 1 \leq 0.$$

We deduce from the Intermediate Value Theorem that there is a number $k \in [0, 1]$ for which $g(k) = 0$. But if $g(k) = 0$, then $f(k) - k = 0$, and so $f(k) = k$.

3.1 (a) Let $\varepsilon > 0$ be given.

We have to find $\delta > 0$ such that

$$|f(x) - f(4)| < \varepsilon \quad \text{whenever} \quad |x - 4| < \delta.$$

Now

$$|f(x) - f(4)| = |(3 - 2x) - (3 - 2 \times 4)| = |8 - 2x|$$
$$= 2|4 - x| = 2|x - 4|.$$

So, when $|x - 4| < \delta$,

$$|f(x) - f(4)| = 2|x - 4| < 2\delta.$$

Therefore, we take $\delta = \varepsilon/2$, and so $|f(x) - f(4)| < \varepsilon$ whenever $|x - 4| < \delta$. Hence f is continuous at 4.

(b) Let $\varepsilon > 0$ be given.

We have to find $\delta > 0$ such that

$$|f(x) - f(a)| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta.$$

Now

$$|f(x) - f(a)| = |(3 - 2x) - (3 - 2a)|$$
$$= |2a - 2x| = 2|a - x| = 2|x - a|.$$

So, when $|x - a| < \delta$,

$$|f(x) - f(a)| = 2|x - a| < 2\delta.$$

Therefore, we take $\delta = \varepsilon/2$, and so $|f(x) - f(a)| < \varepsilon$ whenever $|x - a| < \delta$. Hence f is continuous at a.

3.2 (a) Let $\varepsilon > 0$ be given.

We have to find $\delta > 0$ such that

$$|f(x) - f(4)| < \varepsilon \quad \text{whenever} \quad |x - 4| < \delta.$$

Now

$$|f(x) - f(4)| = |(3x^2 + 2) - (3 \times 4^2 + 2)|$$
$$= |3x^2 - 48| = 3|x^2 - 16|$$
$$= 3|(x - 4)(x + 4)|.$$

So, when $|x - 4| < \delta$,

$$|f(x) - f(4)| = 3|(x - 4)(x + 4)| < 3\delta|x + 4|.$$

Using the Triangle Inequality, we have

$$|x + 4| = |(x - 4) + 8)| \leq |x - 4| + |8| < \delta + 8.$$

Hence, when $|x - 4| < \delta$,

$$|f(x) - f(4)| < 3\delta(\delta + 8).$$

Assume that $\delta \leq 1$; then $3\delta(\delta + 8) \leq 3\delta(1 + 8) = 27\delta$.

We can now make $3\delta(\delta + 8) \leq \varepsilon$ by taking $\delta \leq \varepsilon/27$.

Hence, if $\delta = \min\{1, \varepsilon/27\}$, then $|f(x) - f(4)| < \varepsilon$ whenever $|x - 4| < \delta$.

Thus f is continuous at 4.

(b) To prove that f is continuous on $\mathbb{R}$, we follow the same procedure but use a general point a.

Let $\varepsilon > 0$ be given.

We have to find $\delta > 0$ such that

$$|f(x) - f(a)| < \varepsilon \quad \text{whenever} \quad |x - a| < \delta.$$

Now

$$|f(x) - f(a)| = |(3x^2 + 2) - (3a^2 + 2)|$$
$$= |3x^2 - 3a^2| = 3|x^2 - a^2|$$
$$= 3|(x - a)(x + a)|.$$

So, when $|x - a| < \delta$,

$$|f(x) - f(a)| = 3|(x - a)(x + a)| < 3\delta|x + a|.$$

Using the Triangle Inequality, we have

$$|x + a| = |(x - a) + 2a)| \leq |x - a| + |2a| < \delta + 2|a|.$$

Hence, when $|x - a| < \delta$,

$$|f(x) - f(a)| < 3\delta(\delta + 2|a|).$$

Assume that $\delta \leq 1$; then

$$3\delta(\delta + 2|a|) < 3\delta(1 + 2|a|).$$

We can now make $3\delta(\delta + 2|a|) \leq \varepsilon$ by taking $\delta \leq \varepsilon/3(1 + 2|a|)$.

Hence, if $\delta = \min\left\{1, \dfrac{\varepsilon}{3(1 + 2|a|)}\right\}$, then $|f(x) - f(a)| < \varepsilon$ whenever $|x - a| < \delta$.

Thus f is continuous at a. Since a is an arbitrary point of $\mathbb{R}$, f is continuous on $\mathbb{R}$.

3.3 We must show that, for each positive number ε, there is a positive number δ such that

$$|f(x) - f(0)| < \varepsilon \quad \text{whenever} \quad |x - 0| < \delta.$$

Since $f(0) = 0$, this is equivalent to

$$|f(x)| < \varepsilon \quad \text{whenever} \quad |x| < \delta.$$

We note that $|x| < \delta$ is the same as $-\delta < x < \delta$.

Note that we cannot ensure that x lies in the same part of the domain as 0, so there are two cases to consider:

- when $-\delta < x \leq 0$, and $f(x) = 0$;
- when $0 < x < \delta$, and $f(x) = \sqrt{x}$.

In either case, $|f(x)| \leq \sqrt{|x|}$.

Hence, when $|x - 0| = |x| < \delta$,

$$|f(x) - f(0)| = |f(x)| \leq \sqrt{|x|} < \sqrt{\delta}.$$

Thus, to ensure that $|f(x) - f(0)| < \varepsilon$, we choose $\delta = \varepsilon^2$. Hence $|f(x) - f(0)| < \varepsilon$ whenever $|x - 0| < \delta$. We conclude that f is continuous at 0.

4.1 (a) $f(1, -2) = 20 - (1^2 + (-2)^2) = 15$;

$$g(-1, 3) = \frac{(-1)^2 \times 3}{(-1)^2 + 3^2 + 1} = \frac{3}{11} = 0.2727\ldots.$$

The approximate locations of these points on the surfaces for f and g are shown at the top of the next column.

(b) When $x_1 = 0$, $g(x_1, x_2) = g(0, x_2) = 0$. These points lie on the x_2-axis.

4.2 Let $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$. Then

$$\mathbf{a} - \mathbf{b} = (a_1 - b_1, a_2 - b_2),$$
$$\mathbf{b} - \mathbf{a} = (b_1 - a_1, b_2 - a_2).$$

Thus,

$$d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = \sqrt{(0 - (a_1 - b_1))^2 + (0 - (a_2 - b_2))^2} = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

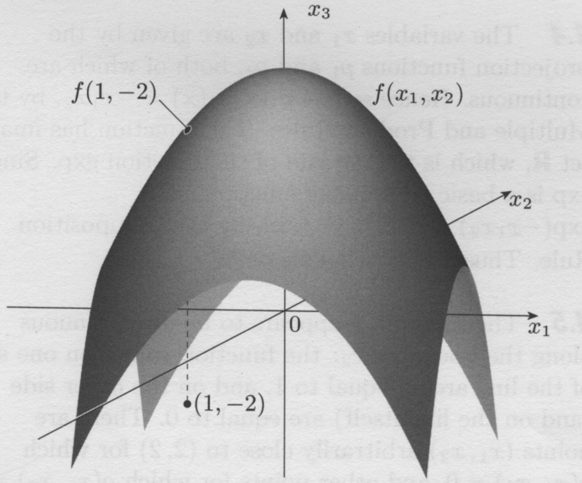
Similarly,

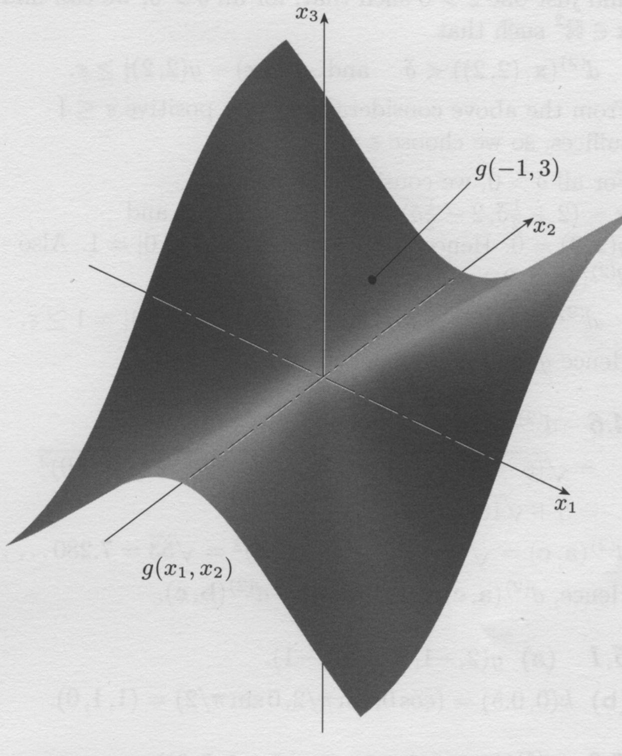
$$d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0}) = \sqrt{(0 - (b_1 - a_1))^2 + (0 - (b_2 - a_2))^2} = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Also

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Thus, $d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0})$.





4.3 The function f consists of adding the variables x_1 and x_2 and then taking the sine of the result. The variables x_1 and x_2 are given by the projection functions $p_1(\mathbf{x}) = x_1$ and $p_2(\mathbf{x}) = x_2$, both of which are continuous. By the Sum Rule, so is $p_1(\mathbf{x}) + p_2(\mathbf{x}) = x_1 + x_2$. This function has image set $\mathbb{R}$, which is the domain of the sine function. The sine function is a basic continuous function. Therefore, by the Composition Rule, we can compose it with $p_1(\mathbf{x}) + p_2(\mathbf{x})$ to give the continuous function $\sin(p_1(\mathbf{x}) + p_2(\mathbf{x})) = \sin(x_1 + x_2) = f(\mathbf{x})$. Thus f is continuous on $\mathbb{R}^2$.

4.4 The variables x_1 and x_2 are given by the projection functions p_1 and p_2, both of which are continuous. Hence so is $-p_1(\mathbf{x})p_2(\mathbf{x}) = -x_1x_2$, by the Multiple and Product Rules. This function has image set $\mathbb{R}$, which is the domain of the function exp. Since exp is a basic continuous function, so is $\exp(-x_1x_2) = e^{-x_1x_2} = f(\mathbf{x})$, by the Composition Rule. Thus f is continuous on $\mathbb{R}^2$.

4.5 The function g appears to be discontinuous along the line $x_1 = x_2$: the function values on one side of the line are all equal to 1, and on the other side (and on the line itself) are equal to 0. There are points (x_1, x_2) arbitrarily close to $(2,2)$ for which $g(x_1, x_2) = 0$, and other points for which $g(x_1, x_2) = 1$.

To prove that g is discontinuous at $(2,2)$, we need to find just one $\varepsilon > 0$ such that, for all $\delta > 0$, we can find $\mathbf{x} \in \mathbb{R}^2$ such that

$$d^{(2)}(\mathbf{x}, (2,2)) < \delta \quad \text{and} \quad |g(\mathbf{x}) - g(2,2)| \geq \varepsilon.$$

From the above considerations, any positive $\varepsilon \leq 1$ suffices, so we choose $\varepsilon = \frac{1}{2}$.

For all $\delta > 0$, we consider the point $\mathbf{x} = (2 + \frac{1}{2}\delta, 2 - \frac{1}{2}\delta)$. We have $g(\mathbf{x}) = 1$ and $g(2,2) = 0$. Hence $|g(\mathbf{x}) - g(2,2)| = |1 - 0| = 1$. Also $d^{(2)}(\mathbf{x}, (2,2)) = \delta/\sqrt{2} < \delta$. Thus

$$d^{(2)}(\mathbf{x}, (2,2)) < \delta \quad \text{and} \quad |g(\mathbf{x}) - g(2,2)| = 1 \geq \varepsilon.$$

Hence g is discontinuous at $(2,2)$.

4.6 $d^{(2)}(\mathbf{a}, \mathbf{b}) + d^{(2)}(\mathbf{b}, \mathbf{c})$

$$= \sqrt{(6-3)^2 + (0-4)^2} + \sqrt{(5-6)^2 + (-3-0)^2}$$
$$= 5 + \sqrt{10} = 8.162\ldots;$$

$d^{(2)}(\mathbf{a}, \mathbf{c}) = \sqrt{(5-3)^2 + (-3-4)^2} = \sqrt{53} = 7.280\ldots.$

Hence, $d^{(2)}(\mathbf{a}, \mathbf{c}) \leq d^{(2)}(\mathbf{a}, \mathbf{b}) + d^{(2)}(\mathbf{b}, \mathbf{c})$.

5.1 **(a)** $g(2, -1, 0) = (0, -1)$.

(b) $k(0, 0.5) = (\cos 0, \sin \pi/2, 0 \sin \pi/2) = (1, 1, 0)$.

5.2 $d^{(5)}((4,2,3,0,-1), (2,-3,-1,0,2))$

$$= \sqrt{(-2)^2 + (-5)^2 + (-4)^2 + 0^2 + 3^2}$$
$$= \sqrt{4 + 25 + 16 + 0 + 9} = \sqrt{54} = 7.348\ldots.$$

5.3 Let $\mathbf{a}$ be a general point in $\mathbb{R}^n$ and let $\varepsilon > 0$ be given. For $\mathbf{x} \in \mathbb{R}^n$,

$$d^{(m)}(f(\mathbf{x}), f(\mathbf{a})) = d^{(m)}(\mathbf{b}, \mathbf{b}) = 0.$$

Hence, for *any* choice of $\delta > 0$,

$$d^{(m)}(f(\mathbf{x}), f(\mathbf{a})) = 0 < \varepsilon \quad \text{whenever} \quad d^{(n)}(\mathbf{x}, \mathbf{a}) < \delta.$$

Hence f is continuous at $\mathbf{a}$. Since $\mathbf{a}$ is an arbitrary point of $\mathbb{R}^n$, f is continuous on $\mathbb{R}^n$.

5.4 The function g can be written as the composition of two functions which we already know to be continuous. Worked problem 5.2 with $\mathbf{u} = \mathbf{0}$ shows that $\mathbf{x} \mapsto d^{(n)}(\mathbf{x}, \mathbf{0})$ is continuous with image set $[0, \infty)$. The square function is one of the basic continuous functions from $\mathbb{R}$ to $\mathbb{R}$. We deduce, by the Composition Rule, that g is continuous on $\mathbb{R}^n$.

5.5 We consider the composition of g with the three projection functions p_1, p_2 and p_3 from $\mathbb{R}^3$ to $\mathbb{R}$, given by

$$(p_1 \circ f)(x_1, x_2) = \cos 2\pi x_1,$$
$$(p_2 \circ f)(x_1, x_2) = \sin 2\pi x_1,$$
$$(p_3 \circ f)(x_1, x_2) = x_2.$$

If we can show that each of these functions is continuous on A then it follows from Theorem 5.6 that f is continuous on A.

To show that $(p_1 \circ f)(x_1, x_2) = \cos 2\pi x_1$ is continuous on A, we show that the constituent parts are continuous and then combine them using the Combination Rules.

In Worked problem 4.3, we proved the continuity of the function $g : \mathbb{R}^2 \to \mathbb{R}$ given by $g(\mathbf{x}) = u_1x_1 + u_2x_2$, where u_1 and u_2 are fixed. It follows, on taking $u_1 = 2\pi$ and $u_2 = 0$, that the function on A given by $(x_1, x_2) \mapsto 2\pi x_1$ is continuous,.

We are also restricting the functions to the set A, and so need to apply the Restriction Rule for functions from $\mathbb{R}^2$ to $\mathbb{R}$. Then, since $x \mapsto \cos x$ is a basic continuous function from $\mathbb{R}$ to $\mathbb{R}$, we can apply the Composition Rule. This shows that $(x_1, x_2) \mapsto \cos 2\pi x_1$ is continuous.

The proof of continuity for $p_2 \circ f$ is similar, except that we use the basic continuous function $x \mapsto \sin x$.

Finally, we note that $p_3 \circ f$ is the function $g(\mathbf{x}) = u_1x_1 + u_2x_2$ of Worked problem 4.3 with $\mathbf{u} = (0, 1)$ restricted to the set A. Hence, by the Restriction Rule, it is continuous. (Alternatively we could simply note that $p_3 \circ f$ is the projection function $p_2 : \mathbb{R}^2 \to \mathbb{R}$ and hence is continuous.)

Since all three compositions of f with a projection function are continuous on A, we conclude from Theorem 5.5 that f is continuous on A.

Index